STRUCTURAL ANALYSES OF THE RIBOSOME BY HYBRID APPROACH

BY

KWOK YAN CHAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Associate Professor Aleksei Aksimentiev, Chair
Professor Klaus Schulten, Director of Research
Professor Taekjip Ha
Professor Jen-Chieh Peng

Abstract

The ribosome is the molecular machine which reads and translates genetic information into proteins in all living cells. Lack of an atomic-resolution structure of the ribosome in its actual functional states prevents our understanding of the ribosome. A hybrid approach overcomes the challenge by combining experimental data from X-ray crystallography and cryo-electron microscopy with computing, permitting one to resolve atomic-level structures of intermediates of the functional ribosome and, thereby, to advance our understanding of ribosome function and the underlying physical mechanisms. In this thesis works we further developed an existing hybrid approach, namely the molecular dynamics flexible fitting (MDFF) method, and apply it to the ribosome. We improved MDFF in two regards, first by incorporating structural symmetry information into the fitting protocol and second by the use of a so-called implicit solvent model. In pursuit of the needed methodological development we participated in the Cryo-EM Modeling Challenge 2010, competing with the MDFF method against other hybrid methods. Two aspects of ribosomal functions were investigated. First we studied bacterial resistance to the antibiotic tetracycline, a study that involved a detail investigation of processes in the ribosome. Second we employed MDFF and molecular dynamics simulations to characterize the dynamics of a ribosome-bound chaperone, called trigger factor.

To my parents

and my beloved, Angela,

for their love, patience and wholehearted support.

Acknowledgements

First of all, I would like to express my grateful gratitude to Prof. Klaus Schulten for his guidance and for his effort in maintaining an excellent research environment with great resources in the Theoretical and Computational Biophysics Group (TCBG). And I thank all the co-workers and collaborators of my thesis projects, including, but not limited to: Prof. Trevor Sewell, Prof. Joachim Frank, Prof. Roland Beckmann, Jean Watermeyer, Wen Li, Julian Deeng, Birgitta Beatrix, James Gumbart, Ryan McGreevy, Leonardo Trabuco, Eduard Schreiner, David Tanner, Jim Phillips and Wei Han. I would also like to thank Prof. Aleksei Aksimentiev, Prof. Taekjip Ha and Prof. Jen-Chieh Peng for their precious time to serve on my thesis committee. They provided invaluable suggestions on my research during my preliminary exam.

During my stay in the TCBG, I received help and guidance from many colleagues, especially from Leonardo Trabuco, Eduard Schreiner and James Gumbart, who tolerated my ignorance and taught me many research skills with patience. I also like to thank Yanxin Liu for the many fruitful discussions about research and science with him, Ryan McGreevy for helping me with many programming issues, Bo Liu for being my officemate and also ally of many reports and proposals writing over the years, and Wei Han for his indispensable assistance with my last thesis project, an extremely challenging project, which cannot be completed without his help.

Throughout my time at UIUC, I shared a lot of enjoyable moments with my friends here: Tak-kei Lui, Billy Lo, Kenny Lo, Tommy So, Ball Wong, Wah-kai Ngai, Jacky Lie, Joseph Ching, Edwin Leung, Karen Ng, Anthony Ho, Juno Mak, Raymond Wong, Boyce Tsang, AtMa Chan, Ricky Chue, Anthony Fan and Kin Lam. I thank all of them for all the fun we had together, which helped to keep me refreshed when I felt tired or overwhelmed by work. In particular, I am very grateful to have a couple of excellent roommates over the years, including Tak-kei Lui, Billy Lo, Kenny Lo and Kin Lam, who have listened to (and tolerated) many of my stupid stories over the years and given me support and advices in times of trouble.

My mother Yin-kwan and my father Yam-choi have always wanted me to choose a medical career.

Yet despite my selfish choice to be a physicist instead, they still supported me wholeheartedly. I apologize for my selfishness to go overseas for years and feel blessed to have their support. Being the eldest son in the family, I also want to thank my younger sister, Ka-yee and my younger brother, Ho-sang, for taking up my role to take care of our parents while I was gone.

Last but not least, I am deeply grateful to my beloved, Angela Wong. Over the years she shared my joy when I was happy, comforted me when I was down, motivated me when I was lost, and most important of all, she loved and supported me unconditionally. She had to stand the loneliness and the pressure from surroundings arising from a long distance relationship, but she didn't give up on me. She is always by my side, in the highs and lows of my life. My dear, I cannot express my gratitude and my love to you in just a few lines, but I promise that, I will spend the rest of my life to do the same to you.

The projects were supported by grants from the National Institutes of Health (9P41GM104601) and the National Science Foundation (PHY0822613). Computer time at the Texas Advanced Computing Center (TACC) was provided by grant MCA93S028 from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. Blue Waters at the National Center for Supercomputing Applications was employed for the projects, under the support of the National Science Foundation (award number OCI 07-25070) and the state of Illinois.

Table of Contents

Chapter 1

Introduction

In all living cells, proteins are synthesized by a molecular machine called the ribosome which reads the genetic information stored on messenger RNA (mRNA) and translates it into protein chains. The ribosome is composed of over 50 different interacting molecular components, including both proteins and RNAs, which assemble into two subunits, namely the large and the small subunit (Fig. 1.1). The two subunits work cooperatively in the translation process, with the small subunit decoding the genetic code on mRNA and the large subunit elongating the nascent protein chain according to the genetic information read evoking the action of a transfer RNA (tRNA) that carries the next amino acid. Besides the ribosomal proteins and RNAs, the ribosome also recruits various factors and ligands to accomplish different tasks during translation. Moreover, protein synthesis is carried out in a series of stages in which the ribosome adopts different conformations. The variability in constituents and conformations possessed by the ribosome renders it one of the most complex biomolecular systems in living cells. Taking advantage of the conserved difference between bacterial and eukaryotic ribosomes, over 50% of efforts developing antibiotics target bacterial ribosomes [1]. Due to its fundamental role in living cells and its medical importance, a better understanding of ribosomal functions is indispensable.

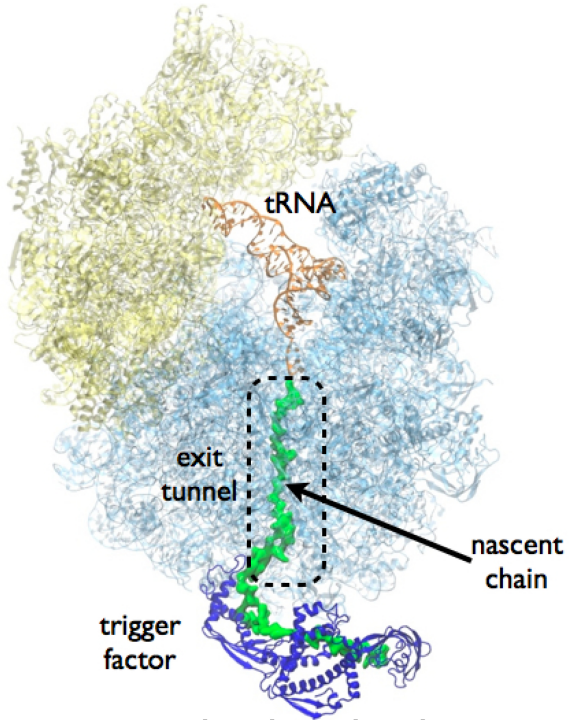

Figure 1.1: Structure of a translating ribosome. The small and large subunits are colored yellow and cyan, respectively, with tRNA colored orange. The elongating nascent chain (green) migrates through the exit tunnel to reach the ribosomal surface. For ribosome synthesizing a cytosolic protein chain, a protein called trigger factor (blue) is recruited to protect the nascent chain from misfolding.

High resolution structural data is crucial to the study of all biomolecular systems, including the ribosome. However, the complexity of ribosomal systems poses a huge challenge to structure determination of the ribosome. An X-ray crystal structure of a full ribosome is available since the last decade. Yet, many of the conformations adopted by the ribosome during elongation of a protein still fail to be crystallized. In particular conformations in which the ribosome forms a complex with factors and ligands are extremely challenging for crystallographic study. Moreover, non-physiological crystalline conditions sometimes render the ribosome non-functional, decreasing the value of such crystal structures. On the other hand, one popular approach to determine structures of the ribosomal complex is cryo-electron microscopy (cryo-EM). In cryo-EM experiments, thousands of samples of the system of interest are frozen instantaneously, keeping the samples in functional form when they are being imaged by the electron microscope. A 3-dimensional structure of the system, in the form of a density map, can be reconstructed from the thousands of images of the samples. Obviating the need of crystallization, cryo-EM has been successfully employed to image ribosomal complexes in different stages of protein synthesis [2]. However, even though EM maps of ribosomal systems at sub-nanometer resolution can be obtained routinely nowadays (5 Å to 10 Å), atomic resolution is still not achieved and direct designation of atomic coordinates into density maps is impossible. Hence, neither the X-ray nor the cryo-EM approach alone can furnish atomic models of various functional forms of the ribosome. Fortunately, a computational method comes to rescue by offering a so-called hybrid approach which bridges the resolution gap between X-ray crystallography and cryo-EM by combining both sources of experimental data to obtain atomic models of ribosomal complexes at various conformations captured by cryo-EM. These atomic models provide the structural information necessary for revealing the mechanisms underlying ribosomal function and for the development of new ribosomal antibiotics to fight bacterial resistance to antibiotics, a major health threat today.

In this thesis we further developed a hybrid method, namely the molecular dynamics flexible fitting (MDFF) method [3, 4] and apply it to the ribosome. First, we incorporated structural symmetry information into the MDFF protocol to enhance the quality of fitted models for symmetric molecular complexes [5]. Second, we showcased the benefits of using a so-called implicit solvent model for MDFF [6]. The performance of MDFF was then evaluated against other hybrid methods in the Cryo-EM Modeling Challenge 2010 [7]. Finally, we applied MDFF to study two ribosomal systems. We employed MDFF to analyze a cryo-EM structure of the ribosome in complex with a ribosomal protection protein Tet(O) and revealed the mechanism of bacterial resistance to the tetracycline antibiotic [8]. We then investigated the dynamics of a ribosome-bound chaperone, called trigger

factor, by MDFF and molecular dynamics simulations [9].

1.1 Molecular dynamics simulations and the molecular dynamics flexible fitting (MDFF) method

The MDFF method [3] is based on MD simulation. In MD simulations, atoms are treated as point particles with different intrinsic properties such as charge, radius, mass, etc. which affect their interactions with the surrounding atoms during the simulations. Interactions among atoms are governed by the potential energy of the system, which usually takes the form of Eq. 1.1 in standard MD force fields. The force field is based on chemical knowledge of inter atomic interactions found in biomolecules. The first three terms describe the interaction energy of atoms involved in chemical bonds and the last two describe the interaction energy of atoms not directly involved in chemical bonds. The bond term refers to energy arising in covalent bonds in the system. The angle term refers to energy arising from angles between pairs of covalent bonds which share an atom as vertex. The dihedral term refers to energy arising from torsion angles of bonds which connect with two other bonds at both ends. The dihedral term also includes the so-called "improper" energy which arises from the planar or non-planar conformation formed by four covalently bonded atoms. The bonded energy terms are illustrated in Fig. 1.2. The non-bonded energy terms include a term for van der Waals interaction, which is represented by Lennard-Jones 6-12 potentials, and a term for Coulomb interactions. Altogether, the potential energy of the biomolecule is given by the expression:

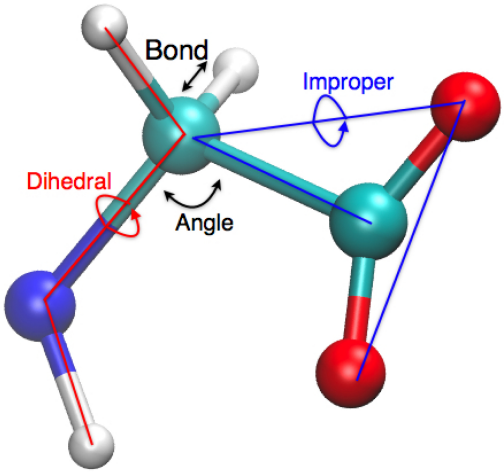

Figure 1.2: Bonded energy terms in MD force field illustrated with a glycine amino acid.

$$
\begin{aligned}
U_{\mathrm{MD}}(\vec{R}) \;=\; & \sum_{bond} k_i^{bond}(r_i - r_0)^2 + \sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2 + \sum_{dihedrals} k_i^{dihedral}[1 + cos(n_i\phi_i + \delta_i)] \\
& + \sum_i \sum_{j>i} 4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}] + \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}
\end{aligned} \tag{1.1}
$$

Here the functional form is chosen to be extremely simple for the sake of fast computation, but this choice comes at the price of limited accuracy.

Using the above potential energy function, the forces acting on atoms can be evaluated. Together with the initial positions and velocities, the trajectories of all atoms and, hence, the dynamics of the simulated system, can be traced. With the advancement in parallel computing and supercomputing facilities, MD simulations of systems as large as a ribosome (about 300,000 atoms) are now achievable, allowing us to study the ribosome by MD simulations as well as applying the MD-based MDFF method.

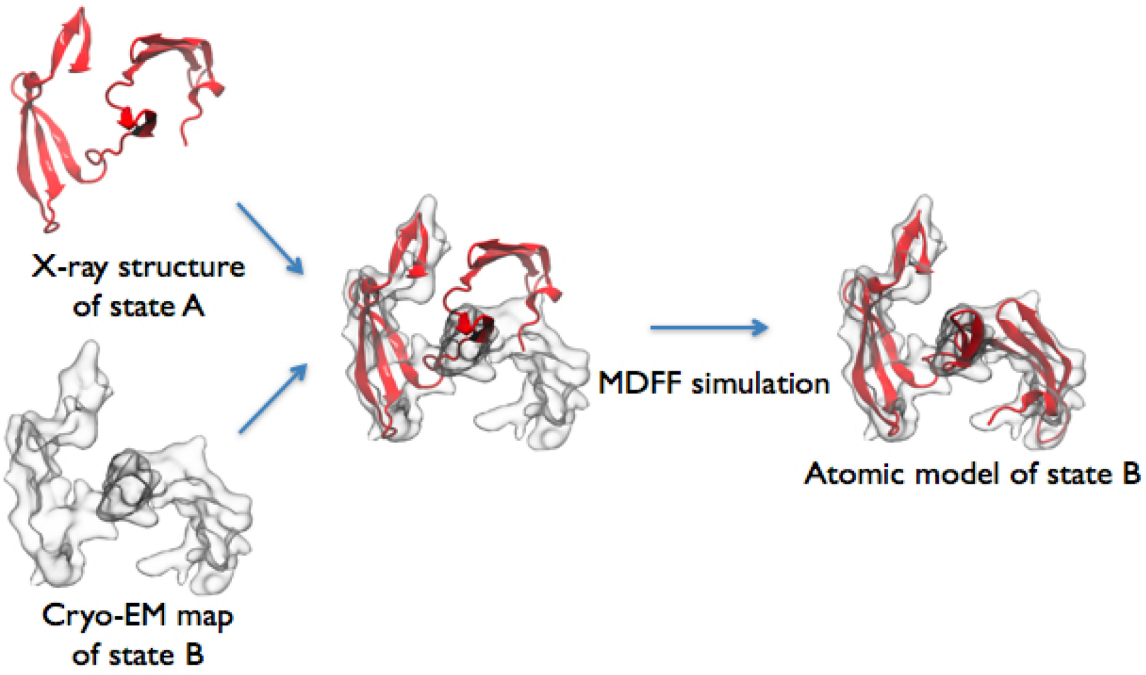

Figure 1.3: Workflow of MDFF.

To employ MDFF, an initial atomic structure, usually an X-ray crystal structure, is subjected to an MD simulation. In this simulation, two additional energy terms are added to the standard MD force field potential U_{MD}. The first term, U_{EM}, is derived from the EM density and is used to apply forces proportional to the gradient of the density map to the atoms using the grid force feature [10] of NAMD [11]. These forces drive the model towards high density regions of the EM map and, hence, effectively guide the structure to a conformational state represented by the EM density (Fig. 1.3). The stereochemical correctness of the structure is ensured by the standard MD force field. In addition, a second term, U_{SS}, which preserves the secondary structure of the system by imposing harmonic restraints on specific dihedral angles and distances between base pairs, is included to avoid overfitting, i.e., to avoid unphysical distortions caused by the U_{EM} term when a structure is too closely

fitted to the EM map. MDFF has already been applied successfully to several ribosomal systems as reported in [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 8].

1.2 Methodological development of MDFF

1.2.1 Symmetry-restrained MDFF

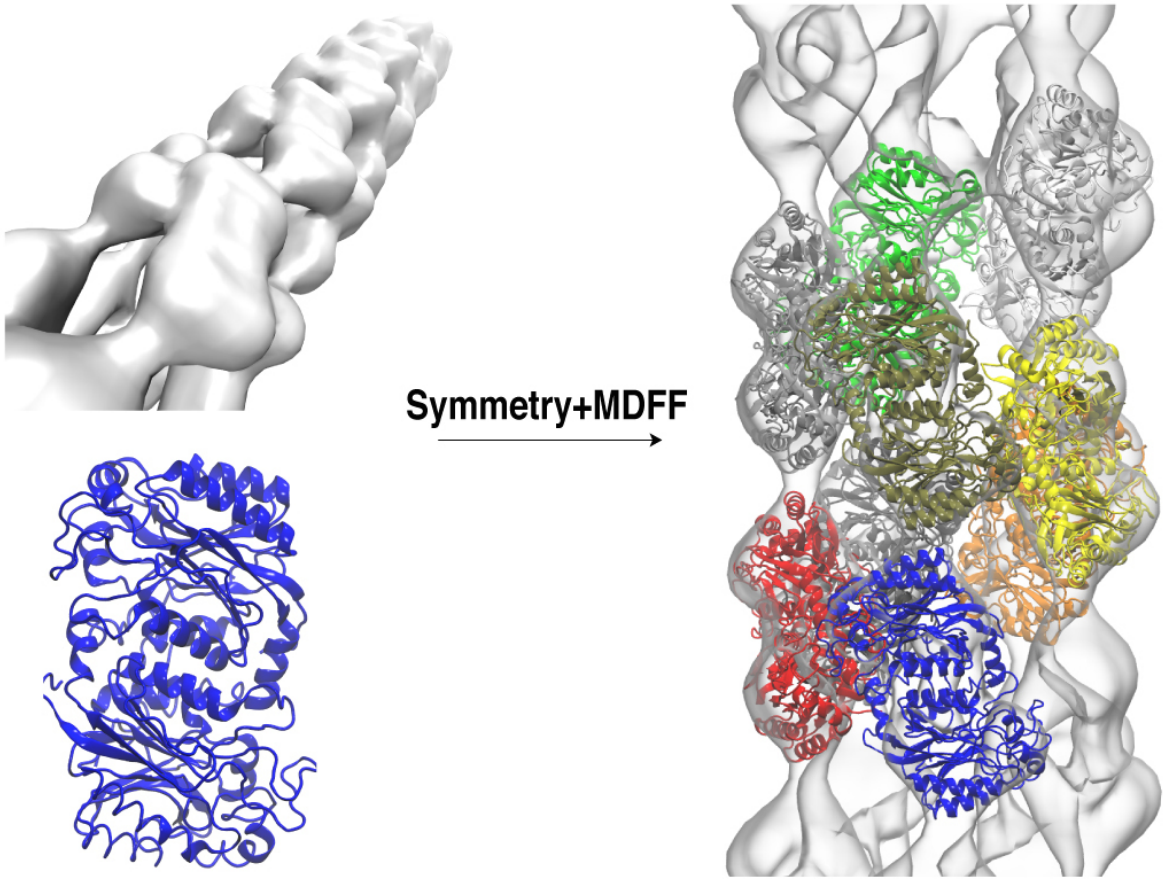

Figure 1.4: Figure illustrating the idea of symmetry-restrained MDFF. Using the symmetry information extracted from EM data (top left), atomic model of a symmetric complex (right) can be built from a single protein unit (bottom left) by symmetry-restrained MDFF. The system shown is a nitrilase from *Rhodococcus rhodochrous* J1 which is a long helical complex.

The quality of MDFF models depends heavily on the resolution of the EM map. A lower resolution map does not represent as much information as a higher resolution map and, hence, the accuracy of a structure fitted to a low-resolution map is limited. Additional knowledge about the structure, such as structural symmetry, can be incorporated into the fitting protocol to reduce the number of available degrees of freedom in the fitting simulations and, hence, improve the quality of fitted models [26, 5]. Indeed, symmetric averaging is often used to improve the signal-to-noise ratio of

an EM dataset [27, 28, 29]. An extension of MDFF, symmetry-restrained MDFF, was developed in which the symmetry information extracted from EM data is incorporated into MDFF (Fig. 1.4). Implementation and benefits of symmetry-restrained MDFF are described in Chapter 2 and were published in [5].

1.2.2 MDFF with implicit solvent model

Similar to normal MD simulations, MDFF requires an accurate solvent model [30, 31] to produce high quality fitted structures. Indeed, many shortcomings of MDFF are largely due to the use of in vacuo simulations, in which the solvent environment is represented by a single dielectric constant, omitting many effects of the solvent. On the other hand, explicit solvent model [32] represents the solvent environment by explicit water molecules, and is regarded as accurate and natural for MD simulations. However, the inclusion of explicit water molecules largely increases the atom count of a simulation, often by a factor of ten, rendering MDFF with explicit solvent model computationally too expensive. The generalized Born implicit solvent (GBIS) model implemented in NAMD [11, 6, 33] offers a great further advance to the MDFF method by furnishing the necessary description of solvent environment while avoiding inclusion of a huge number of water molecules into MDFF simulations. The absence of explicit water molecules also eliminates the viscosity imposed on simulated solutes, effectively allowing faster equilibration of solute conformations and better conformational sampling and, hence, speeding up conformational changes in MDFF simulations. The benefits of employing the GBIS model in MDFF are described in Chapter 3 and were published in [6].

1.2.3 Cryo-EM Modeling Challenge 2010

Besides MDFF, there are many other hybrid approaches which adopt different algorithms to perform flexible fitting. As a first effort to bring cryo-EM modelers together to evaluate the performance of various hybrid approaches in a systematic way, the Cryo-EM Modeling Challenge 2010 was held in which developers of different hybrid methods applied their methods to model the same set of cryo-EM maps and deposited the resultant structures onto a public-accessible server [34]. We participated in the challenge by applying MDFF to obtain atomic models for seven density maps provided by the challenge, covering a wide range of system sizes (single proteins to large complexes) and resolutions (2.5–8 Å). The results of our participation are described in Chapter 4 and were published in [7].

1.3 The ribosomal protection protein Tet(O)

The ribosome is the target of many antibiotic drugs [1]. However, the continuously growing bacterial resistance to antibiotics has rendered many existing antibiotics ineffective, creating a constant need for new types of antibiotics. To design new types of antibiotics, it is indispensable to understand the physical mechanism of bacterial resistance to antibiotics.

Tetracycline (Tc), a previously widely-used antibiotic [35], is nowadays of limited use because of the emergence of bacterial resistance. One resistance mechanism employed by bacteria involves binding of a so-called ribosomal protection protein, Tet(O), to the bacterial ribosome which actively removes Tc from the bacterial ribosome [36, 37]. However, due to the lack of high resolution structural data of Tet(O)-bound ribosome, the mechanism behind Tet(O)-mediated resistance remains elusive. In collaboration with Joachim Frank (Columbia U.), MDFF was applied to a cryo-EM map of Tet(O)-bound ribosome at 9.6 Å to obtain the first atomic model of the system (Fig. 1.5). The model of the ribosome-Tet(O) complex revealed three critical loops of Tet(O) which promote the unbinding of Tc from the ribosome in various ways. Findings from our cryo-EM analyses were confirmed by mutation experiments. The results are described in Chapter 5 and were published in [8].

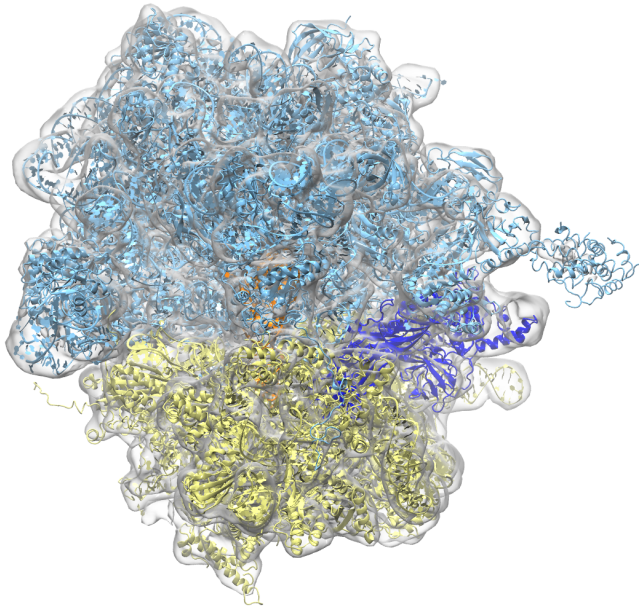

Figure 1.5: MDFF model of a Tet(O)-bound ribosome inside the EM density map. The Tet(O) protein is colored in dark blue.

1.4 The ribosome-bound chaperone trigger factor

During protein synthesis, nascent protein chains exit the ribosome through the ribosomal exit tunnel while still being elongated inside the ribosome (Fig. 1.1). Various types of ligands bind to the ribosome to interact with the exiting nascent chain for different functional purposes. For example, in case of a membrane protein nascent chain, the ribosome docks to a SecY system on a cellular membrane to co-translationally insert the nascent protein into the membrane [21]; in case of cytosolic nascent

proteins, a protein called trigger factor (TF) binds to the ribosomal exit site (Fig. 1.1) to protect the nascent chain from pre-mature folding and aggregation [38, 39] by retarding folding [40, 41] or even by unfolding folded domains of the nascent chain during translation [41]. Deletion of the TF and another chaperone, DnaK, from living cells results in a decreasing yield of *de novo* protein folding and can be lethal [42, 43]. However, the actual chaperone activities of the TF remain elusive.

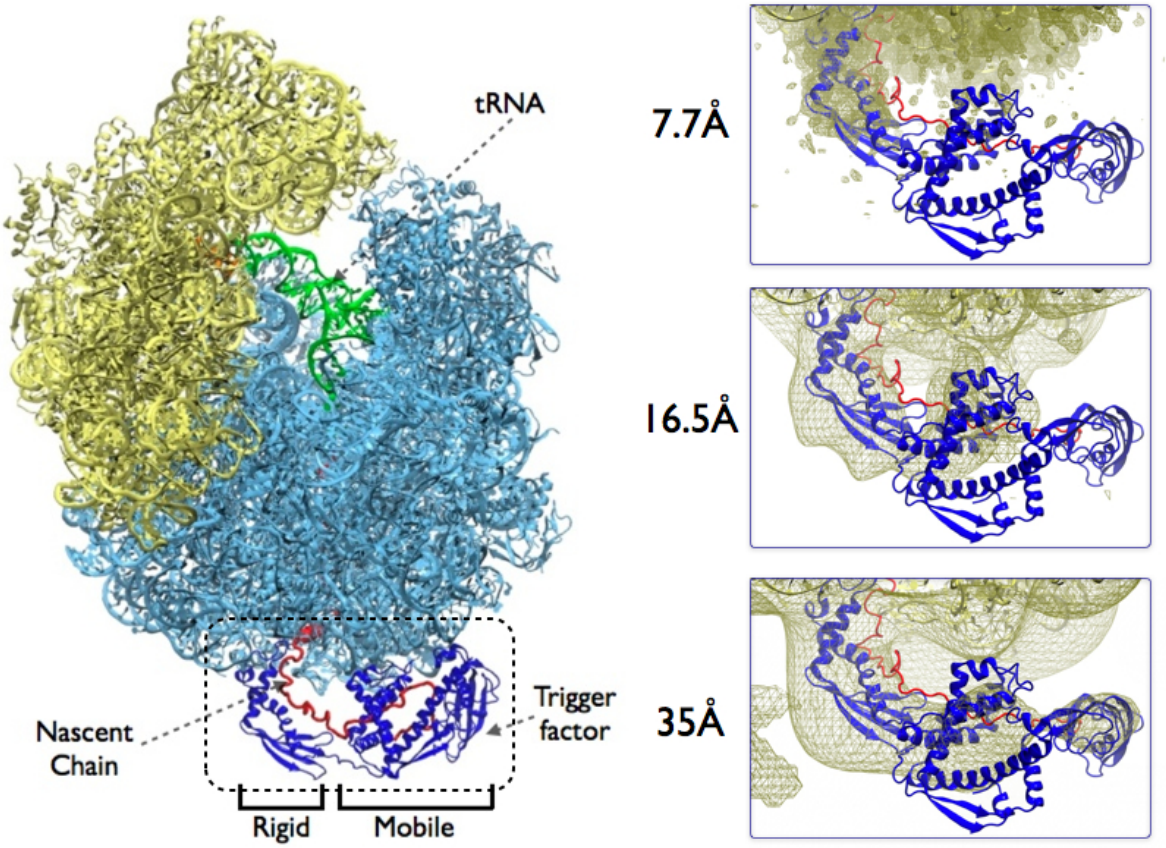

Figure 1.6: The trigger factor is poorly resolved by cryo-EM beyond the ribosome binding domain. Densities of the other two domains are only visible when the map is low-pass filtered to 16.5 Å and 35 Å, respectively, suggesting a highly mobile TF on the ribosome.

To resolve the molecular mechanism underlying the functions of the TF, high resolution structural data of TF binding to a translating ribosome are needed. Our collaborator Roland Beckmann (U. Munich) resolved cryo-EM structures of the ribosome-trigger factor-nascent chain complex. However in the cryo-EM data, two out of three TF domains are poorly resolved, suggesting a highly dynamic TF on the ribosomal surface (Fig. 1.6). We first applied MDFF to obtain an atomic model for regions that are well resolved in the EM map, and then performed microseconds-long simulations to sample the conformations of the poorly resolved TF domains. Our simulations reveal that the dynamics of

TF is dependent on the length of the nascent peptide; this dependence originates from the interactions between the TF and the nascent chain and is physiologically important for the chaperone function of the trigger factor. The results are described in Chapter 6 and were submitted for publication [9].

Chapter 2

Symmetry-restrained flexible fitting

2.1 Introduction

Structural information is essential for understanding function and mechanism of biological systems. The state of the art in macromolecular structure determination permits the structural characterization of such molecules at multiple resolutions. X-ray crystallography, the predominant high-resolution technique, provides structural data at the atomic scale, but often requires non-physiological conditions to achieve crystallization. Furthermore, certain macromolecular assemblies, such as the ribosome or those with quaternary helical structure, are difficult to crystallize or cannot be crystallized at all. On the other hand, cryo-electron microscopy (cryo-EM) can capture images of molecules in functional states and for large systems, but typically at lower resolution (3.3-30 Å) than X-ray crystallography [44, 45, 13, 46, 21].

The resolution gap between X-ray crystallography and cryo-EM can be bridged by using hybrid computational methods to combine the two sources of experimental data. An early approach is the so-called rigid-body docking, which treats X-ray structures as rigid and searches for an orientation that best fits the cryo-EM map [47]. Even though this approach provides an excellent first approximation to the fit, it lacks the flexibility needed to deal with molecules that exhibit relative motions between individual parts in different functional states. Hence, a new class of flexible fitting methods has been developed, in which the X-ray structures or homology models have more degrees of freedom to deform and fit the cryo-EM map. An early attempt at flexible fitting involved dividing the molecule into different parts and then docking each rigid part into the density independently [48]. Another early approach utilized real-space refinement, a technique developed for X-ray crystallography, to fit the

structure into the map [49, 50]. Other means of flexible fitting include (1) matching reduced-complexity representations of the structure and the density map to deform the structure [51], (2) altering the structure along the low frequency normal modes to increase the correspondence to the density map [52, 26, 53], (3) fitting comparative models based on different sequence-structure alignments and loop conformations of different components [54], (4) using a deformable elastic network and restraints from EM data to morph the structure [55], and (5) refining segments of protein homology models locally within EM data before global refinement of the whole protein [56]. A recently developed flexible fitting method, molecular dynamics flexible fitting (MDFF), employs molecular dynamics (MD) simulations to perform flexible fitting by incorporating the EM data as an external potential in conventional MD simulations [3].

In any of these flexible fitting methods, the quality of the final fitted structure is strongly related to the resolution of the map and the quality of the starting model. A lower resolution map does not represent as much information as a higher resolution map and, hence, allows more degrees of freedom to participate a fitting procedure. Moreover, the accuracy of a structure fitted to a low-resolution map maybe limited by structural variability in the dataset, which reduces the useful structural information one can extract. One way to improve the quality and accuracy of structures fitted into EM data, especially in case of low-resolution maps, is to incorporate additional knowledge about the structure in the fitting procedure. For example, many biological systems are symmetric in nature, such as poliovirus exhibiting icosahedral symmetry [57] and potassium channels exhibiting four-fold symmetry [58]. Point group symmetry and helical symmetry also occur frequently in biological systems. This symmetry is functionally important, e.g., for increasing structural stability [59, 60, 61], for cooperative or allosteric functions [62, 63], and for folding [64, 65]. Indeed, symmetry information of biomolecules has already been incorporated in structure prediction tools like Rosetta [66]. Protocols to impose restraints enforcing symmetry during MD simulations have also been developed and have been used in the refinement of homology models of potassium channels [67]. Symmetry information can be extracted even from low-resolution EM data, and symmetric averaging is commonly used to improve the signal-to-noise ratio of a dataset. By integrating the same symmetry information into flexible fitting protocols to ensure a symmetric fitted structure, one can effectively reduce the number of available degrees of freedom in the fitting problem and, hence, improve the quality of fitted models. A similar idea has been employed in normal mode-based fitting of virus capsids in which only modes obeying the icosahedral symmetry were retained [26].

We present here a method to incorporate the symmetry information extracted from EM data into

MDFF. The symmetry relationship among subunits is converted into harmonic restraints, which are then applied to maintain the structure's symmetry during MDFF simulations. The benefit derived from symmetry restraints is most apparent for low-resolution EM data as high-resolution data already contains sufficient information to maintain the symmetry during MDFF simulations. We applied the symmetry restraints to three different systems for which experimental EM data are available, namely the GroEL-GroES complex, a nitrilase from *Rhodococcus rhodochrous* J1, and a chaperonin from the archaeon *Methanococcus maripaludis* (Mm-cpn). These test cases represent data having three different symmetries and a range of resolutions, and serve to illustrate the benefits of applying symmetry restraints during MDFF.

2.2 Results

Symmetry restraints

The MDFF method is based on molecular dynamics simulation [3] and has already been used successfully in numerous applications [17]. Starting from an initial atomic structure, usually an X-ray crystal structure, an MD simulation is performed. In this simulation, two additional energy terms are added to the standard MD force field potential U_{MD}. The first term, U_{EM}, is derived from the EM density, and is used to apply forces proportional to the gradient of the density map to the atoms. This drives the model towards high density regions of the map and, hence, effectively guides the structure to the conformational state represented by the EM density. The stereochemical correctness of the structure is ensured by the standard MD force field. In addition, a second term, U_{SS}, which preserves the secondary structure of the system by imposing harmonic restraints on specific dihedrals, is included to avoid overfitting, i.e., to avoid unphysical distortions caused by the U_{EM} term when a structure is too closely fitted to the EM map. A practical guide for carrying out MDFF simulations has been published previously [4]. Since MDFF is based on MD simulations, thermal fluctuations in the fitting simulations can lead to multiple solutions that fit a map. Hence, different symmetric units can arrive at different conformations even for a symmetric map.

The symmetry information of the cryo-EM data can be incorporated into an MDFF simulation through an additional potential to maintain a symmetric structure during the simulations. Since it is a common practice to symmetrize EM data for symmetric molecules during the 3D EM reconstruction process [27, 28, 29], the type of symmetry and its parameters, e.g., helical rise and twist for a helically symmetric system, can be obtained from this process.

A new potential energy term U_{SR} is built according to the symmetry information as follows. Let $\mathbf{R}_i(t)$ be the set of coordinates for atoms of the i-th symmetric unit at time t during the simulation, where $\mathbf{R}_i(t)$ for different units are related by a predefined symmetry. The symmetry information is used to generate, for each unit i, a transformation $\overleftrightarrow{U}_i$, which superimposes the coordinates $\mathbf{R}_i$ of unit i onto the first unit. Averaging over these superimposed symmetric units, the coordinates of an average structure, $\mathbf{R}_{avg}(t) = <\overleftrightarrow{U}_i\ \mathbf{R}_i(t)>$, can be calculated. Using the inverse transformation $\overleftrightarrow{U}_i^{-1}$ for each unit, the average structure is transformed backwards from the first unit to the respective i-th unit; the resulting set of backward-transformed coordinates, $\mathbf{R}'_i(t) = \overleftrightarrow{U}_i^{-1}\ \mathbf{R}_{avg}(t) = \overleftrightarrow{U}_i^{-1} <\overleftrightarrow{U}_i\ \mathbf{R}_i(t)>$ are now perfectly symmetric. The deviation between $\mathbf{R}'_i(t)$ and $\mathbf{R}_i(t)$ is measured by their root mean square distance ($RMSD_i$), defined by

$$RMSD_i(t) = \sqrt{< |\mathbf{R}_i(t) - \mathbf{R}'_i(t)|^2 >} = \sqrt{< |\mathbf{R}_i(t) - \overleftrightarrow{U}_i^{-1} <\overleftrightarrow{U}_i\ \mathbf{R}_i(t)> |^2 >} \tag{2.1}$$

where the average is taken over a chosen subset of atoms in the i-th unit. A high $RMSD_i$ indicates a large deviation from the ideal symmetric structure.

To minimize $RMSD_i$ between each unit and its corresponding unit in the average structure, the potential energy term U_{SR} mentioned above is added, given by

$$U_{\mathrm{SR}} = \frac{1}{2} k(t) \sum_i [RMSD_i(t)]^2 \tag{2.2}$$

where the sum is taken over all units. This potential term defines forces applied to the chosen subset of atoms in order to minimize $RMSD_i(t)$ during the simulations and, hence, guide the system towards a symmetric structure. The factor $k(t)$ with unit energy/length^2 controls the strength of the applied forces and can either be set to a constant or be linearly increased over time. A $k(t)$, linearly increasing up to a finite value, is recommended so that during the early steps of fitting, the structure has sufficient flexibility to explore the conformational space represented by the EM data before converging to a symmetric structure as the symmetry restraint forces increase. The subset of atoms within a unit on which the RMSD calculation is based and to which forces are applied is typically chosen to be the C_α atoms for proteins instead of all atoms, so that the structure will not be over-restrained.

We applied MDFF with and without symmetry restraints to three exemplary applications, namely GroEL-GroES complexes, nitrilase from *R. rhodochrous* J1 and Mm-cpn (see Figure 2.1). For all three

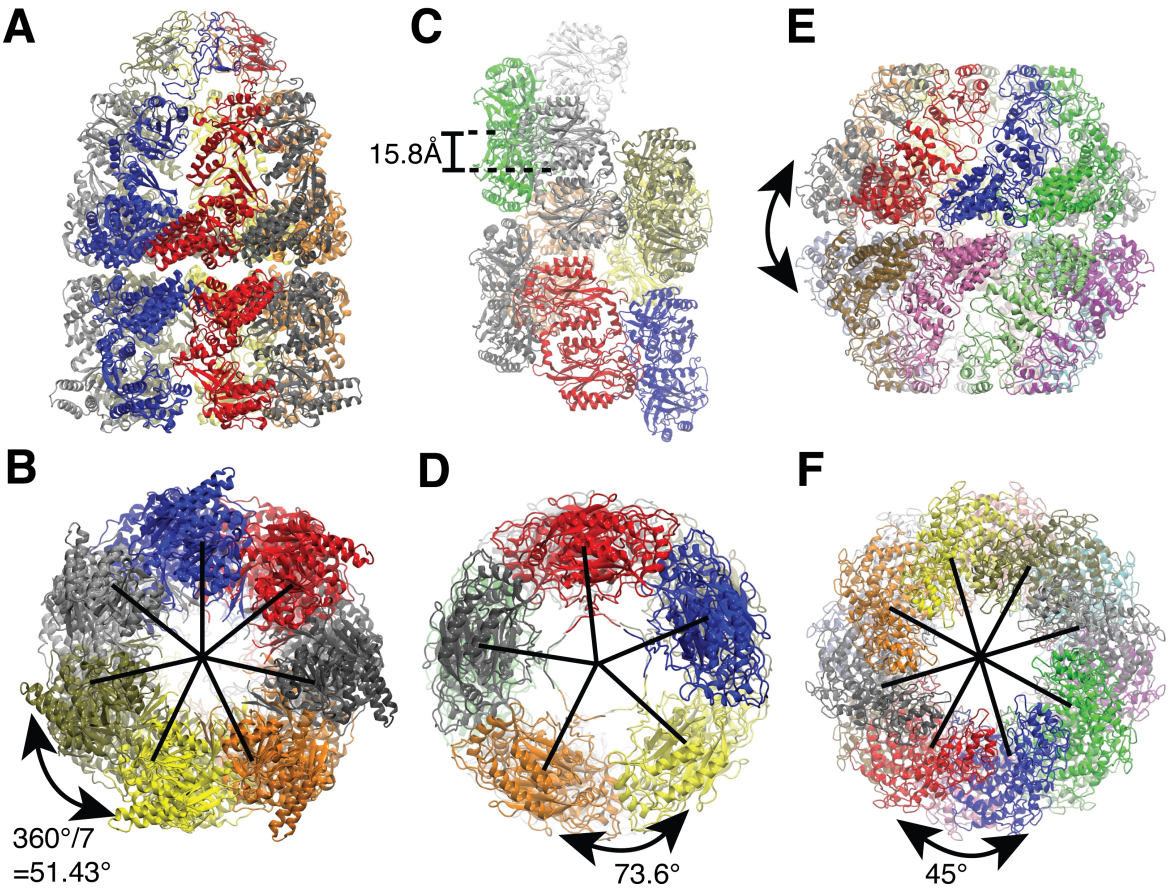

Figure 2.1: Symmetries of the example applications, showing side and end-on views of each. (A, B) Seven-fold symmetric GroEL-GroES complex (C_7 symmetry). (C, D) Helically symmetric nitrilase of *R. rhodochrous* J1 ($D_1S_{4.9}$ symmetry). (E, F) Sixteen-fold symmetric units of Mm-cpn (D_8 symmetry).

cases, we calculated along the fitting trajectory the average RMSD between symmetric units, defined by $< RMSD_i(t) >= \frac{1}{N}\sum_i RMSD_i(t)$, where N is the number of symmetric units. As shown in Figures 2.2A,B and D, in all three examples, the average RMSD between symmetric units for fitting with symmetry restraints applied is lower than that for fitting without symmetry restraints. The lower RMSD demonstrates that the symmetry restraints work as intended.

Example 1: GroEL-GroES complexes

The first example involves MDFF for the GroEL-GroES complex. The *Escherichia coli* chaperonin GroEL, together with the lid-like co-chaperonin GroES, form a molecular machine that assists the folding of many proteins with the help of ATP binding and hydrolysis. Upon ATP hydrolysis, the GroEL-GroES complex undergoes conformational changes necessary to carry out its function of mediating protein folding [68]. Crystal structures of an analogue of the ATP-bound state [69] and a

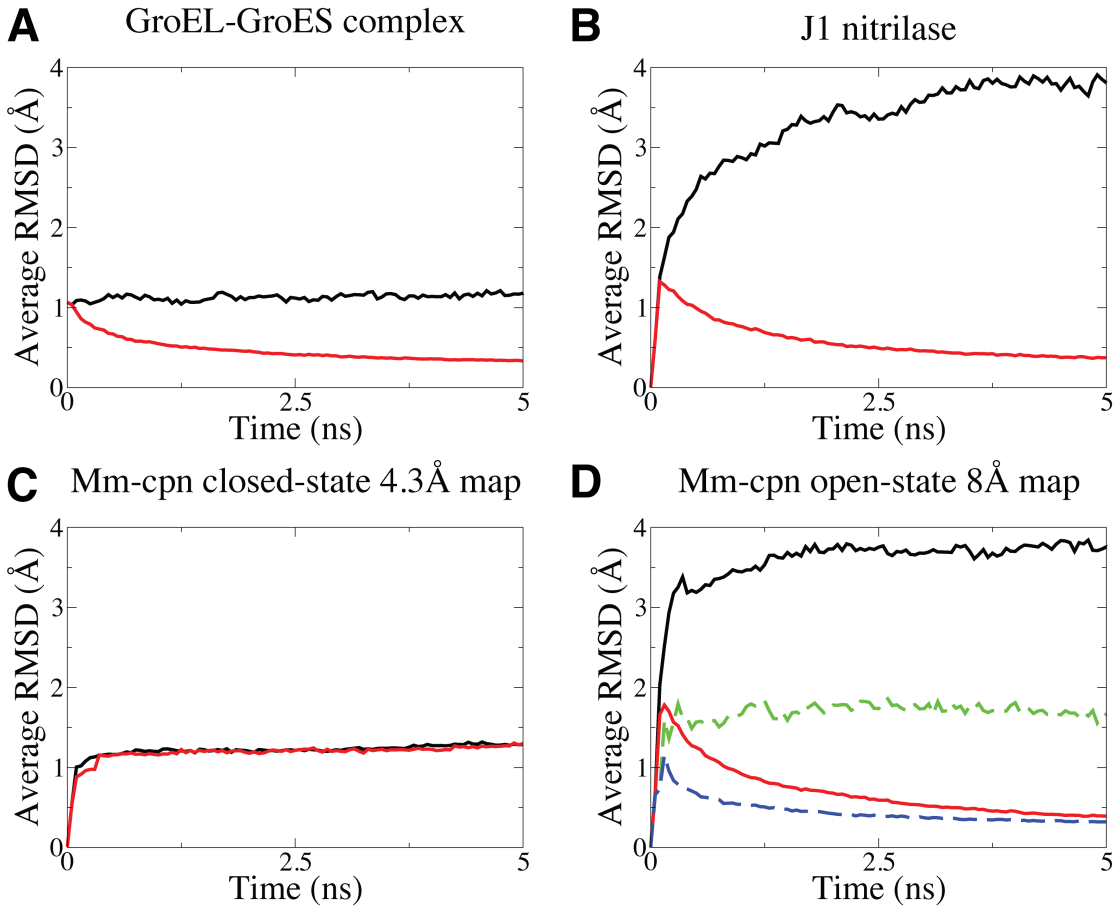

Figure 2.2: Average C_α RMSD values along the fitting trajectories of (A) GroEL-GroES complex, (B) Nitrilase of *R. rhodochrous* J1, (C) Mm-cpn closed state (4.3 Å) and (D) Mm-cpn open state (8 Å). The average RMSD value is defined through $< RMSD_i(t) >= \frac{1}{N}\sum_i RMSD_i(t)$ where N is the number of symmetric units. Red and black curves in (A)-(D) correspond to fitting with and without symmetry restraints, respectively. In (D), solid lines represent fitting using a homology model, while dashed lines represent fitting using a crystal structure with (blue) and without (green) symmetry restraints.

post-hydrolysis ADP-bound state [70] are available, but appear to be identical, failing to resolve the structural and functional differences between the two states. In contrast, both states have been captured in cryo-EM maps that have successfully resolved potentially important structural differences between them [71], suggesting that crystal packing may have favored non-physiological conformations in the published crystal structures.

GroEL comprises two back-to-back seven-membered rings and, together with the co-chaperonin GroES, forms a complex with a seven-fold rotational symmetry, as illustrated in Figures 2.1A and B. We applied MDFF to fit an ADP-bound structure (PDB code 2C7D), which was modelled from cryo-EM data [71], into the EM map of an ATP-bound complex, both with and without symmetry restraints. The target cryo-EM data used here is an intermediate resolution (7.7 Å) map of the

complex in an ATP-bound state [71], obtained from the EM data bank (EMD-1180).

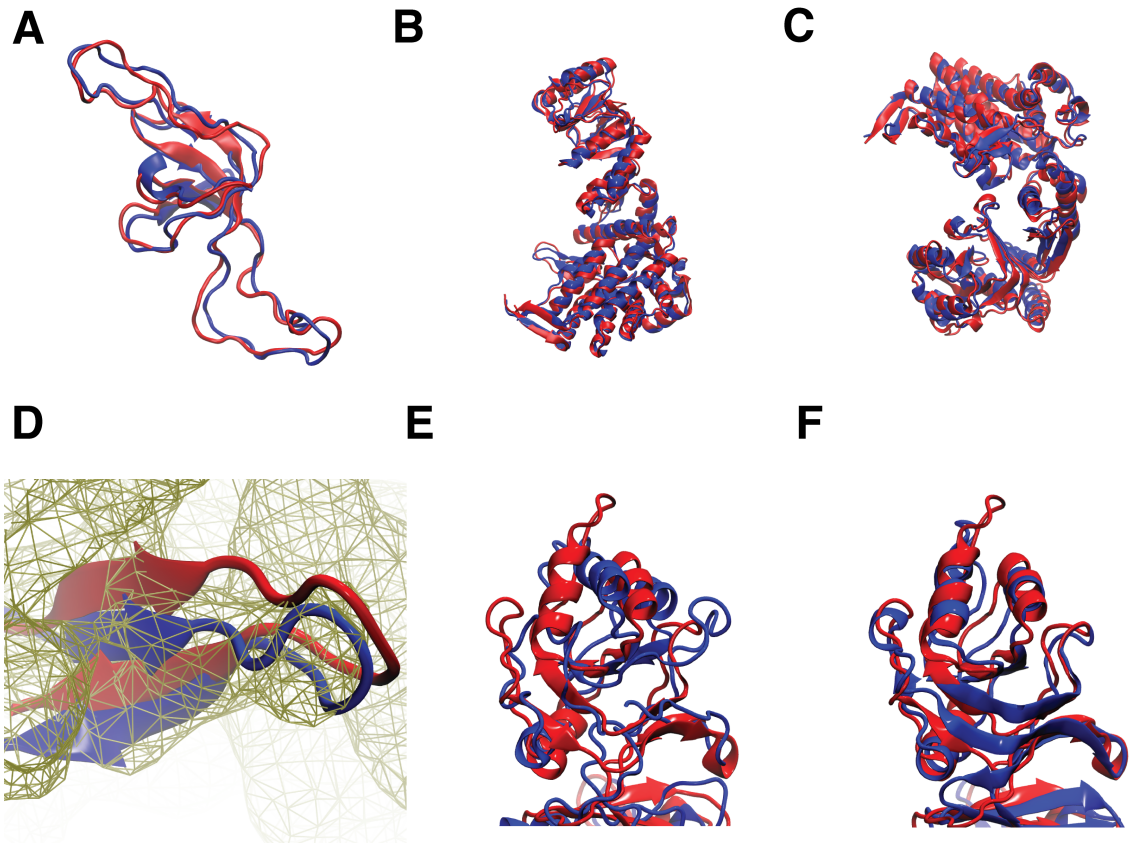

Figure 2.3: Comparison of symmetry-restrained fitted structures (blue) to known structures (red) of the corresponding proteins. (A) GroES (RMSD 1.73 Å). (B) GroEL *cis*-ring (RMSD 1.15 Å). (C) GroEL *trans*-ring (RMSD 2.51 Å). (D) β-sheet region (residues 37-51 of the GroEL *trans*-ring subunit) of the symmetry-restrained fitted structure (blue) and the published structure (red) in the ATP-bound state inside the EM map. (E,F) Apical domain of the Mm-cpn lidless open state. Both a homology model (E, RMSD 5.31 Å) and a crystal structure (F, RMSD 2.34 Å) of a closed state were used for fitting.

To validate the structure fitted with symmetry restraints, a comparison was made to an available structure (PDB code 2C7C) modelled from the same map used here. The RMSD of the full GroEL-GroES complex, compared to the published structure, decreases from 2.21 Å to 1.97 Å (fitted without symmetry restraints) and 1.93 Å (fitted with symmetry restraints) during the fitting. From Figure 2.3 it can be seen that GroES (RMSD 1.73 Å, Figure 2.3A), the GroEL *cis*-ring (RMSD 1.15 Å, Figure 2.3B) and the GroEL *trans*-ring (RMSD 2.51 Å, Figure 2.3C) of the fitted structure are all very similar to the EM-modelled structure. Furthermore, the chief observable structural difference between the published ATP- and ADP-bound structures modelled from EM data, namely a shorter inter-strand distance between β-sheets in adjacent subunits of the *trans*-ring in the ATP-bound state [71], is also captured in our fitted structure. However, the separation between the β-sheets observed in our fitted structure of the ATP-bound state is wider. The wider inter-strand distance agrees better with the

map than the published structure, as shown in Figure 2.3D. Such a difference can be attributed to the fact that the *trans*-ring of the published structure was obtained by rigid-body docking of three domains to the map, which failed to capture intra-domain motions.

After fitting, free equilibration of both the symmetry-restrained and unrestrained fitted structures was carried out. From the 10-ns trajectory of both simulations, the backbone RMSD of the entire complex with respect to the first frame (i.e., the final fitted structure) was calculated. In Figure 2.4A, it can be seen that the simulation starting from a structure fitted with symmetry restraints equilibrates faster and has a lower backbone RMSD than the one starting from a structure fitted without symmetry restraints. The shorter time needed to equilibrate and the lower RMSD value (Figure 2.4A) demonstrate that a more stable structure has been obtained using the symmetry restraints during the fitting. The stability of the fitted structure is particularly important if the system is to be subjected to further MD simulations.

Example 2: Nitrilase

The second example involves a nitrilase from *R. rhodochrous* J1, a member of the superfamily of nitrilases, amidases, acyl transferase and *N*-carbamoyl-D-amino acid amidohydrolases. Nitrilases convert nitriles to the corresponding carboxylic acids and ammonia; an oligomerization of individual protein dimers into spiral homo-oligomers is important for their enzymatic function [72, 73]. In particular, the nitrilase from *R. rhodochrous* J1 was found to be inactive in its dimer form, but active in its helical-fiber form [74]. Negative stain and cryo-EM have been used to resolve the 3D structure of these helical fibers for native nitrilases as well as mutants [73, 75], in order to understand the relationship between function and spiral quaternary structure formation in the nitrilase family.

We applied MDFF to fit a two-turn-helix model of *R. rhodochrous* J1 nitrilase, consisting of nine dimers, into a low resolution (18 Å) negative stain EM map of a long helix fiber (EMD-1313) [74]. The atomic structure of the dimer of *R. rhodochrous* J1 nitrilase used is a homology model based on crystal structures of four nitrilase homologues (PDB codes: 2PLQ, 3HKX, 1J31, 2VHH (Thuku, personal communication)). The helical symmetry of the fiber is parametrized by the helical twist (azimuthal rotation around the helical axis) and the helical rise (rise along the helical axis) between adjacent dimers. The helical twist and rise of the published map were -73.65° and 15.8 Å, respectively, with a $D_1S_{4.9}$ symmetry [74].

Because only two helical turns were used for the atomic model, while the map covered six turns, the first and the last dimers at the edge of the model were attracted by the adjacent, empty density.

Indeed, in the fitting without symmetry restraints, certain parts of the first and last dimers were pulled into this density, leading to structural distortions, as illustrated in the blue structure in Figure 2.4B. However, in the fitting with symmetry restraints, both the first and last dimers maintained their dimer form inside their corresponding density envelope (red structure in Figure 2.4B). The difference in dimer behavior shows that the symmetry restraints prevent edge-distortion effects, which arise whenever one is fitting only a part of the full system into the map. Segmenting the maps and removing the density where no dimers are placed may be a possible way to avoid edge-distortion effects, but map segmentation is prone to human error and proper cutting of boundaries is not always achieved. Although this problem can be addressed more directly by doing a fitting for the full system, there are occasions in which fitting of only a portion is desirable. In the J1 nitrilase system, a two-turn helix model is sufficient for determining how inter-dimer interactions give rise to the stability of a spiral structure; fitting more dimers into the long helical EM data would be computationally costly and unnecessary. This strategy may be appropriate to other symmetries such as virus capsids, which have icosahedral symmetry and for which full-system fitting would be computationally prohibitive.

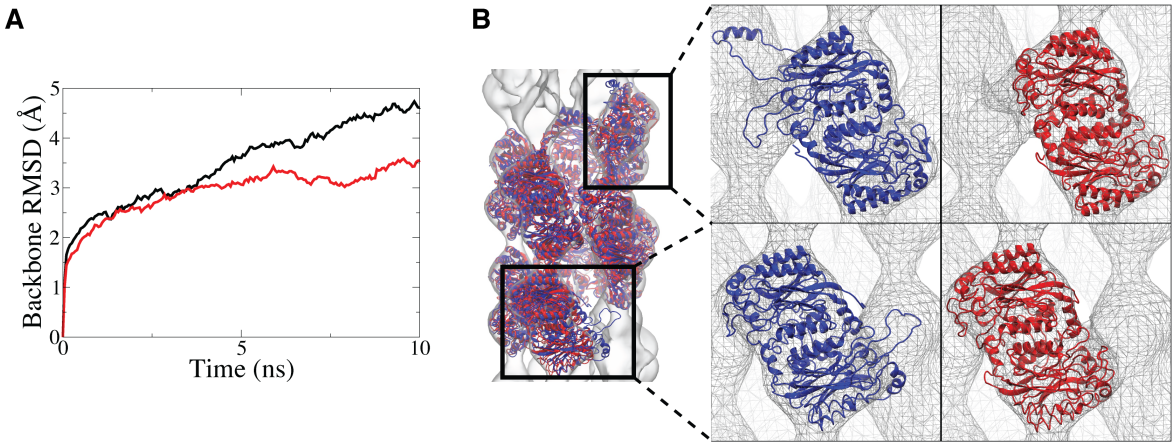

Figure 2.4: (A) Backbone RMSD of the GroEL-GroES complex with respect to the fitted structure during free equilibration after fitting. Red and black curves correspond to starting structures which were fitted with and without symmetry restraints, respectively. (B) Two-turn-helix fitted structure of *R. rhodochrous* J1 nitrilase. Red and blue structures were fitted with and without symmetry restraints, respectively. The first and last dimers of the structure fitted without symmetry restraints are pulled towards the density of adjacent dimers and are distorted as a consequence, while these dimers remain intact in case of structures fitted with symmetry restraints.

Example 3: Mm-cpn

The last example features Mm-cpn, an archaeal group II chaperonin, which mediates protein folding. Mm-cpn is composed of two back-to-back eight-membered rings, but unlike the GroEL-GroES complexes, the rings are related by a two-fold axis giving rise to 16 symmetric units in the system.

To carry out its function, a built-in lid closes upon ATP hydrolysis, the latter inducing structural rearrangements which have been illuminated by cryo-EM [76].

We applied MDFF to model the structures represented by two distinct maps, one at medium resolution (8 Å) of a lidless Mm-cpn in the open state (EMD-5140) and one at high resolution (4.3 Å) of a full Mm-cpn in the closed state (EMD-5137) [76]. For the medium-resolution lidless open-state map, MDFF was used to fit a lidless closed-state homology model, built from the sequence of a lidless Mm-cpn and a full closed-state structure (PDB code 3LOS), into the EM data. For the high-resolution full closed-state map, MDFF was used to refine and improve the full closed-state structure (PDB code 3LOS).

The Mm-cpn example showcases the usefulness of symmetry restraints in dealing with lower (8 Å) resolution EM data. As shown in Figure 2.2C, fitting with and without symmetry restraints gives a similar average RMSD for C_α atoms among the symmetric units in the 4.3-Å-resolution map. However, for the lower resolution (8 Å) open-state map, the symmetry restraints are required to preserve the symmetry of the system. For high-resolution data, the conformational space allowed by the map is more confined than for low-resolution maps and, hence, during the fitting to them, all units converge to the same conformation in the absence of symmetry restraints. In contrast, there are significantly more possible conformations for a symmetric unit fitted to a lower resolution map, causing different units' conformations to diverge during the fitting and, thus, giving a non-symmetric structure when fitted without the symmetry restraints.

We compared our symmetry-restrained fitted model of the Mm-cpn lidless open state to a recently released crystal structure (PDB code 3KFK [77]), finding that the apical domain exhibits noticeable structural differences between them (RMSD 5.32 Å, see Figure 2.3E). Such differences may be due to the inaccuracy of the initial homology model of a lidless closed-state. To eliminate this source of inaccuracy, we performed fittings for the lidless open state map using a different starting structure, a crystal structure of the lidless closed state (PDB code 3KFE [77]). As shown in Figure 2.3F, a structure in better agreement with the open-state crystal structure (RMSD 2.34 Å) was obtained. The improvement in the fitted model demonstrates the importance of using a good starting structure for fitting. Furthermore, the structure obtained with symmetry-restrained fitting is in better agreement with the crystal structure (RMSD 2.34 Å) compared to fitting without symmetry restraints (RMSD 2.83 Å), showing again that symmetry-restrained fitting can improve the quality of fitted structures.

2.3 Discussion

We have developed a new type of restraint that maintains the symmetry of a system during MDFF simulations by utilizing symmetry information extracted from cryo-EM data. Application to Mm-cpn demonstrates that when employing MDFF without symmetry restraints, the symmetry is distorted more heavily during fitting of lower resolution data, namely 8 Å vs. 4.3 Å data, because for each symmetric unit, more structural variability is allowed by a lower resolution map. Hence, symmetry restraints should be employed when one is modeling low to medium resolution symmetric EM data by MDFF. However, even when using symmetry restraints, the quality of the starting structure plays a significant role in the fitting, exemplified by Mm-cpn (see Figures 2.3E,F).

As shown in the free equilibration simulations of the GroEL-GroES complex, a more stable structure, i.e., one more suitable for further MD simulations, can be obtained with the application of symmetry restraints during fitting. The finding is in agreement with previous simulations of homology models of potassium channels refined by symmetry-restrained MD simulations [67], which is an example of an alternative use of symmetry restraints in MD simulations other than MDFF. Another benefit of using symmetry restraints is the prevention of edge-distortion effects, which arise when one is fitting only a portion of the system into the map, as illustrated above for *R. rhodochrous* J1. The benefits of using symmetry restraints demonstrated for the three example systems are also applicable to biological complexes possessing other types of symmetry, such as the icosahedral symmetry of virus capsids. The generalization is straightforward as the transformations arising in Equations 2.1 and 2.2 can be defined also for other symmetry types.

In symmetry-restrained fitting simulations, thermal noise tends to drive the system away from symmetry whereas the restraints counteract this tendency, limiting conformations to those that remain close to the symmetric structure. At physiological temperature, small deviations from symmetry arise naturally, especially at the side-chain level. Hence, in general it is not recommended to apply symmetry restraints to all atoms, which can produce a structure that adheres to the symmetry perfectly, but may also be non-physiological due to overly restraining it, akin to the over-fitting problem for cryo-EM maps [3]. Instead, applying symmetry restraints to C_α atoms only can avoid over-restraining while still maintaining the secondary and tertiary structural symmetry of the system. The balance between thermal fluctuations and symmetry is controlled by the force constant $k(t)$, which is recommended to be increased linearly over the fitting simulation so that the system can explore more conformational space as allowed by the map before converging to a symmetric structure. A maximum force constant

of 10 (kcal/mol)/Å^2 was chosen in our examples, from which deviations of atomic positions from the average will be $\Delta x \sim \sqrt{k_B T/k} = 0.25$ Å only, thus maintaining the symmetry well. Indeed, in our examples, a linearly increasing $k(t)$, reaching 10 (kcal/mol)/Å^2, proved to be sufficient to guide the system to a nearly symmetric structure (see convergence of average RMSD in Figures 2.2) within a simulation time of 5 ns.

There are alternatives to symmetry restraints that can give a symmetric fitted structure. For example, one may fit only one symmetric unit into a segmented map and then generate the whole structure using the symmetry transformations. Although this will result in a perfectly symmetric structure, fitting in the presence of other symmetric units should be preferred as this will produce a more reliable structure at the interfaces between different units. Indeed, clashes may arise when one attempts to generate the whole structure from only a single fitted subunit, while in symmetry-restrained fitting the MD force field prevents such clashes. One may also average the individual structures of all symmetric units at the end of the fitting protocol, performing a so-called "instantaneous symmetrization". This will give a perfectly symmetric structure, but the average of units in different conformations is not likely to be physically realistic or biological relevant. In contrast, the symmetry restraints will bias all units towards the same conformation during the fitting. The advantage of using symmetry restraints during the simulations over "instantaneous symmetrization" has also been demonstrated previously using symmetry-restrained MD simulations of tetrameric potassium channels [67].

Symmetry is common and often functionally important for biological systems, but there are cases where symmetry breaking arises during function. An example occurs during viral entry into the cell, which starts with binding of the icosahedrally symmetric virus capsid to receptors on the cell membrane, inducing then a conformational transition of the capsid for injection of the genome into the cell [78]; clearly, upon binding to the receptor, the symmetry of the virus capsid is broken. Naturally, symmetry-restrained fitting does not capture highly asymmetric regions in a system given that the map itself is symmetric, but subsequent MD simulations without any restraints applied permit the development of asymmetry. Root-mean-square fluctuations (RMSF) of atoms during simulations, which measure the deviation of atoms from their average positions, can reveal regions that are more flexible and, hence, more likely to become asymmetric.

It is common for multiple symmetries to co-exist in a biological system; for example, the family of nitrilases exhibits a two-fold symmetry in addition to the helical symmetry. Therefore, extending symmetry restraints to handle multiple symmetries in a single simulation is desirable and, indeed, is already developed. A tutorial on applying symmetry restraints during an MDFF simulation has been

developed and is available at http://www.ks.uiuc.edu/Training/Tutorials/.

2.4 Methods

All systems utilized in the example applications were first rigid-body docked into the EM map using Colores from the Situs package [51], and then solvated in a box of TIP3P [32] water molecules, using 10 Å padding in all directions. Counter ions Na^+ and Cl^- were added to neutralize the systems. The simulations were performed with a development version of NAMD 2.8 [11] using the CHARMM27 force field with CMAP corrections [79, 80]. Only water and ions were allowed to equilibrate for the first 500 ps by constraining the protein with harmonic restraints, followed by equilibration of side chains as the protein backbone remained constrained. Next, MDFF simulations were performed for 5 ns in each case, coupling only non-hydrogen atoms of the protein to the U_{EM} potential with a grid scaling of 0.3, which controls the balance between U_{MD} and U_{EM} [3]. In addition to secondary structure restraints, restraints were used to maintain the correct chirality at all chiral centers and to keep peptide bonds in the trans-configuration. Finally 3000 steps of energy minimization, in the presence of U_{EM} with grid scaling of 10, were performed to increase the stability of the resulting structure by removing the thermal deviations in the systems. For simulations using symmetry restraints, the force constant k was increased linearly from 0 to 10 (kcal/mol)/Å^2 during the 5-ns MDFF simulations and forces were applied only to C_α atoms. All simulations were carried out in the NVT ensemble, using the following parameters: constant temperature at 300 K was maintained using Langevin dynamics with a damping constant of 5 ps^{-1}; long-range electrostatic forces were computed using the particle-mesh Ewald summation method with a grid spacing of 1 Å; the RESPA multiple-time-stepping algorithm [81, 82] was employed with an integration time step of 1 fs, short-range forces evaluated every 2 time steps, and long-range electrostatics evaluated every 4 time steps.

Chapter 3

MDFF with implicit solvent model

3.1 Introduction

Molecular dynamics (MD) is a computational method [83] employed for studying the dynamics of nanoscale biological systems on nanosecond to microsecond timescales [84]. Using MD, researchers can utilize experimental data from crystallography and cryo-electron microscopy (cryo-EM) to explore the functional dynamics of biological systems [4].

Because biological processes take place in the aqueous environment of the cell, a critical component of any biological MD simulation is the solvent model employed [30, 31]. An accurate solvent model must reproduce water's effect on solutes such as the free energy of solvation, dielectric screening of solute electrostatic interactions, hydrogen bonding and van der Waals interactions with solute. For typical biological MD simulations, solute is comprised of proteins, nucleic acids, lipids or other small molecules.

Two main categories of solvent models are explicit and implicit solvents. Explicit solvents, such as SPC [85] and TIP3P [32], represent water molecules explicitly as a collection of charged interacting atoms and calculate a simple potential function, such as Coulomb electrostatics, between solvent and solute atoms. Implicit solvent models, instead, ignore atomic details of solvent and represent the presence of water indirectly through complex interatomic potentials between solute atoms only [86, 87, 88]. There are advantages and disadvantages of each solvent model.

Simulation of explicit water is both accurate and natural for MD, but often computationally too demanding, not only since the inclusion of explicit water atoms increases a simulation's computational cost through the higher atom count, but also because water slows down association and disassociation

processes due to the relatively long relaxation times of interstitial water [89]. The viscous drag of explicit water also retards large conformation changes of macromolecules [90].

An alternative representation of water is furnished by implicit solvent descriptions which eliminate the need for explicit solvent molecules. Implicit water remains always equilibrated to the solute. The absence of explicit water molecules also eliminates the viscosity imposed on simulated solutes, allowing faster equilibration of solute conformations and better conformational sampling. Examples of popular implicit solvent models are Poisson-Boltzmann electrostatics [91, 92], screened Coulomb potential [87, 93], analytical continuum electrostatics [94] and generalized Born implicit solvent [95].

The generalized Born implicit solvent (GBIS) model, used by MD programs CHARMM [96, 97], Gromacs [98, 99], Amber [100] and NAMD [11, 101], furnishes a fast approximation for calculating the electrostatic interaction between atoms in a dielectric environment described by the Poisson-Boltzmann equation. The GBIS electrostatics calculation determines first the Born radius of each atom, which quantifies an atom's exposure to solvent, and, therefore, its dielectric screening from other atoms. The solvent exposure represented by Born radii can be calculated with varying speeds and accuracies [102] either by integration over the molecule's interior volume [103, 104] or by pairwise overlap of atomic surface areas [95]. GBIS calculations then determine the electrostatic interaction between atoms based on their separation and Born radii.

GBIS has benefited MD simulations of small molecules [105]. For the case of large systems, whose large conformational motions [106] may benefit most from an implicit solvent description, but which must be simulated on large parallel computers [107], requiring efficient parallel GBIS algorithms. NAMD addresses the computational challenges of parallel GBIS calculations and efficiently simulates large systems. Details of NAMD's GBIS model and the implementation are described in [6].

The molecular dynamics flexible fitting (MDFF) method [3, 4, 10] is a MD simulation method that matches crystallographic structures to an electron microscopy (EM) map; crystallographic structures often correspond to non-physiological states of biopolymers while EM maps correspond often to functional intermediates of biopolymers. The MDFF method adds to a conventional MD simulation an EM map-derived potential, thereby driving a crystallographic structure towards the conformational state represented by an EM map. Shortcomings of MDFF are largely due to the use of in vacuo simulations; such use was necessary hitherto as simulations in explicit solvent proved too cumbersome. Implicit solvent MDFF simulations promise a significant improvement of the MDFF method. To demonstrate the benefit of using NAMD's GBIS model in MDFF, we simulate the ∼250,000-atom ribosome using MDFF with different solvent models and compare behavior of different solvent models in these MDFF

simulations.

3.2 Results

The ribosome is the cellular machine that translates genetic information on mRNA into protein chains. During translation, tRNAs, with their anti-codon loops to be matched to the genetic code on mRNA, carry amino acids to the ribosome. The synthesized protein chain is elongated by one amino acid each time a cognate tRNA (with its anti-codon loop complementary to the next mRNA codon) brings an amino acid to the ribosome; a peptide bond is formed between the new amino acid and the existing protein chain. The ribosome complex fluctuates between two conformational states, namely the so-called classical and ratcheted state [108], during the elongation process. The transition from classical to ratcheted state involves multiple, large conformational changes, including an inter-subunit rotation between its 50S and 30S subunits [108] and the closing of its L1 stalk in the 50S subunit [109] (see Figure 3.1). The large conformational changes during the transition from classical to ratcheted state are essential for translation [110] as suggested by previous cryo-EM data [45]. MDFF-derived models of the classical and ratcheted state ribosome provide atomic-level details crucial to understanding protein elongation in the ribosome.

A high-resolution classical state ribosome structure was fitted into a low-resolution ratcheted state EM map in an in vacuo MDFF simulation as well as MDFF simulations employing explicit and implicit solvent. The rate of convergence and relative accuracy of solvent models in the three MDFF simulations are characterized by the root-mean-square deviation between models, defined as follow:

$$\mathrm{RMSD}_{\mathrm{sol,ref}}(t) = \sqrt{\sum_{i}^{N} \left[\vec{r}_{i,\mathrm{sol}}(t) - \vec{r}_{i,\mathrm{ref}}\right]^2 / N} \,, \tag{3.1}$$

where $\vec{r}_{i,\mathrm{sol}}(t)$ denotes the atomic coordinates at time t of the simulation corresponding to one of the three solvent models (exp, imp or vac) and $\vec{r}_{i,\mathrm{ref}}$ denotes the atomic coordinates for the last time step ($t_f = 3$ ns) of the simulation using the reference solvent model (exp, imp or vac) as specified below. Unless otherwise specified, the summation is over the $N = 146,000$ heavy atoms excluding the mRNA, L10 and L12 protein segments which are too flexible to be resolved by the cryo-EM method.

Figure 3.1 plots $\mathrm{RMSD}_{\mathrm{exp,exp}}(t)$, $\mathrm{RMSD}_{\mathrm{imp,exp}}(t)$ and $\mathrm{RMSD}_{\mathrm{vac,exp}}(t)$ that compare each MDFF simulation against the final structure reached in the explicit solvent case. We note that using the initial rather than final structure as the reference could yield a slightly different characterization of

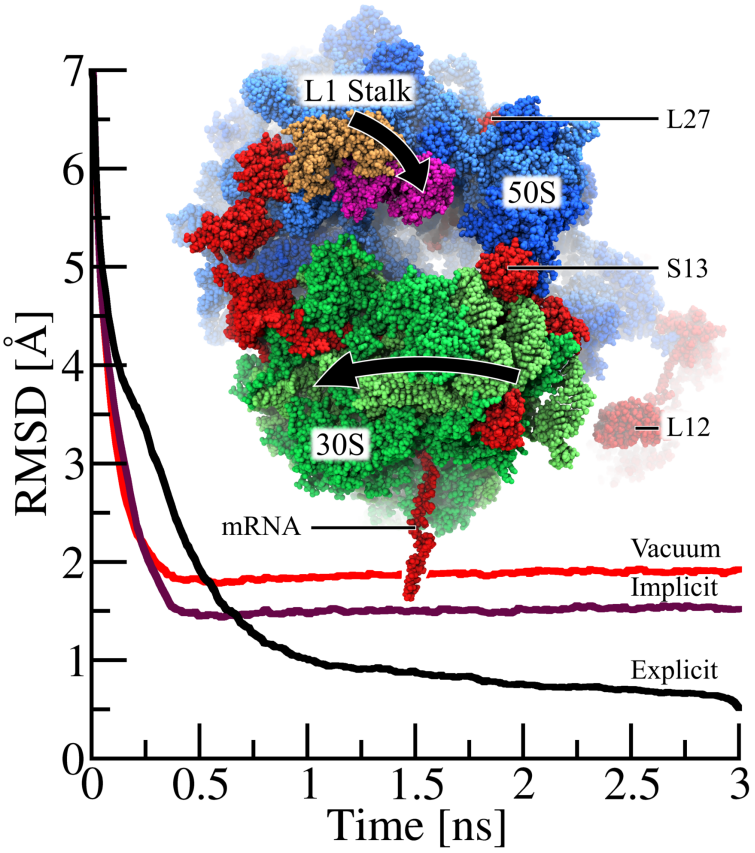

Figure 3.1: Molecular dynamics flexible fitting (MDFF) of ribosome with NAMD's GBIS method. While matching the 250,000-atom classical ribosome structure into the EM map of a ratcheted ribosome, the 30S subunit (green) rotates relative to the 50S subunit (blue) and the L1 stalk moves 30 Å from its classical (tan) to its ratcheted (magenta) position. Highlighted (red) are regions where the implicit solvent structure agrees with the explicit solvent structure much more closely than does the in vacuo structure. The root-mean-squared deviation ($\mathrm{RMSD}_{\mathrm{sol,exp}}(t)$) of the ribosome, defined in eq. 3.1, with the final fitted explicit solvent structure as reference, is plotted over time for explicit solvent ($\mathrm{RMSD}_{\mathrm{exp,exp}}(t)$ in black), implicit solvent ($\mathrm{RMSD}_{\mathrm{imp,exp}}(t)$ in purple) and in vacuo ($\mathrm{RMSD}_{\mathrm{vac,exp}}(t)$ in red) MDFF. While the explicit solvent MDFF calculation requires 2-3 ns to converge to its final structure, both implicit solvent and vacuum MDFF calculation require only 0.5 ns to converge. As seen by the lower RMSD values for $t > 0.5$ ns, the structure derived from the implicit solvent fitting agrees more closely with the final explicit solvent structure than does the in vacuo structure. While this plot illustrates only the overall improvement of the implicit solvent structure over the in vacuo structure, the text discusses key examples of ribosomal proteins (L27, S13 and L12) whose structural quality is significantly improved by the use of implicit solvent.

convergence [111], e.g., a slightly different convergence time. As manifested by $\mathrm{RMSD}_{\mathrm{imp,exp}}(t)$ and $\mathrm{RMSD}_{\mathrm{vac,exp}}(t)$, the implicit solvent and vacuum MDFF calculations converge to their respective final structures in 0.5 ns compared to 2-3 ns for the explicit solvent case, i.e., for $\mathrm{RMSD}_{\mathrm{exp,exp}}(t)$.

The final structures obtained from the MDFF simulations are compared in Table 3.1 through the $\mathrm{RMSD}_{\mathrm{sol,ref}}(t)$ values for $t = 3$ ns. The ribosome structure from GBIS MDFF closely agrees with the one from explicit solvent MDFF as indicated by the value $\mathrm{RMSD}_{\mathrm{imp,exp}}(3\,\mathrm{ns}) = 1.5$ Å; the in vacuo MDFF ribosome structure, however, compares less favorably with the explicit solvent MDFF structure as suggested by the larger value $\mathrm{RMSD}_{\mathrm{vac,exp}}(3\,\mathrm{ns}) = 1.9$ Å. While the 0.4 Å improvement

in RMSD of the GBIS MDFF, over in vacuo MDFF, structure implies an overall enhanced quality, certain regions of the ribosome are particularly improved.

		Reference		
		exp	imp	vac
Solvent	exp	0	1.5	1.9
	imp	1.5	0	2.1
	vac	1.9	2.1	0

Table 3.1: Root-mean-square deviation ($\mathrm{RMSD}_{\mathrm{sol,ref}}(3\,\mathrm{ns})$ in Å) between the three final ribosome structures matched using explicit solvent, GBIS, and in vacuo MDFF. The GBIS and explicit solvent MDFF structures closely agree as seen by $\mathrm{RMSD}_{\mathrm{imp,exp}}(3\,\mathrm{ns}) = 1.5$ Å, while the in vacuo MDFF structure deviates from the explicit solvent MDFF structure by $\mathrm{RMSD}_{\mathrm{vac,exp}}(3\,\mathrm{ns}) = 1.9$ Å. See also Figure 3.1.

The regions with the highest structural improvement (highlighted red in Figure 3.1) belong to segments at the exterior of the ribosome and to segments not resolved by and, therefore, not coupled to the EM map, i.e., not being directly shaped by MDFF. For proteins at the exterior of the ribosome, GBIS MDFF produces higher quality structures than in vacuo MDFF, because these proteins are highly exposed to solvent and, therefore, require a solvent description. The structural improvement for several exterior solvated proteins, calculated by $\mathrm{RMSD}_{\mathrm{vac,exp}}(3\,\mathrm{ns}) - \mathrm{RMSD}_{\mathrm{imp,exp}}(3\,\mathrm{ns})$, is 3.5 Å , 2.4 Å and 1.6 Å for ribosomal proteins S6, L27 and S13 (highlighted red in Figure 3.1), respectively. Accurate modeling of these proteins is critical for studying the translation process of the ribosome. The L27 protein, for example, not only facilitates the assembly of the 50S subunit, it also ensures proper positioning of the new amino acid for peptide bond formation [112]. The S13 protein, located at the interface between subunits, is critical to the control of mRNA and tRNA translocation within the ribosome [113].

The use of GBIS for MDFF also increases structural quality in regions where the EM map does not resolve the ribosome's structure and, therefore, MDFF does not directly influence conformation; though it is most important that MDFF correctly models structural regions defined in the EM map, it is also desirable that it correctly describes regions of crystal structures not resolved by the EM map. The structural improvement, over in vacuo MDFF, of the unresolved segments is 8.3 Å for mRNA and 4.9 Å for L12 (highlighted red in Figure 3.1). The L12 segment is a highly mobile ribosomal protein in the 50S subunit that promotes binding of factors which stabilize the ratcheted conformation; L12 also promotes GTP hydrolysis which leads to mRNA translocation [114]. As clearly demonstrated, the use of GBIS MDFF, instead of in vacuo MDFF, improves the MDFF method's accuracy for matching crystallographic structures to EM maps, particularly for highly solvated or unresolved proteins.

To compare computational performance of the solvent models for MDFF, each ribosome simulation was benchmarked on 1020 processor cores (3.5 GHz processors with 5 GB/s network interconnect); the simulation speed for explicit solvent MDFF is 3.6 ns/day, for implicit solvent MDFF it is 5.2 ns/day and for vacuum MDFF it is 37 ns/day. GBIS MDFF performs 50% faster than does explicit solvent MDFF, but seven times slower than in vacuo MDFF. NAMD's GBIS implementation is clearly able to achieve a more accurate MDFF match of the ribosome structure (see Table 3.1) than does an in vacuo MDFF calculation and does so at a lower computational cost than explicit solvent MDFF.

3.3 Discussion

The generalized Born implicit solvent (GBIS) model has long been employed for molecular dynamics simulations of relatively small bio-molecules. NAMD's unique GBIS implementation can also simulate very large systems, such as the entire ribosome, and does so efficiently on large parallel computers. The new GBIS capability of NAMD offers a great further advance to the MDFF method by furnishing the necessary description of solvent environment while avoiding inclusion of a huge number of water molecules into MDFF simulations. Moreover, the absence of explicit water molecules eliminates the viscosity imposed on simulated solutes, effectively allowing faster equilibration of solute conformations and better conformational sampling and, hence, speeding up conformational changes in MDFF simulations.

3.4 Methods

Generalized Born implicit solvent model

A theoretical description of the generalized Born implicit solvent model is provided in Appendix A.

MDFF simulations of the ribosome

The classical state in our simulations is an all-atom ribosome structure [16] with 50S and 30S subunits taken from PDB IDs 2I2V and 2I2U, respectively [115], and the complex fitted to an 8.9 Å resolution classical state EM map [45]. In the multistep protocol for fitting this classical state ribosome to a ratcheted state map [3], the actual ribosome is fitted first, followed by fitting the tRNAs. Since the fitting of the ribosome itself exhibits the largest conformational changes (inter-subunit rotation and L1-stalk closing), we limit our MDFF calculation here to the ribosome and do not include tRNAs.

Three MDFF simulations were performed using NAMD [11] and analyzed using VMD [116]. The MDFF simulations are carried out in explicit TIP3P [32] solvent, in implicit solvent and in vacuo. All simulations were performed in the NVT ensemble with the AMBER99 force field [117], employing the SB [118] and BSC0 [119] corrections and accounting for modified ribonucleosides [120]. The grid scaling parameter [4], which controls the balance between MD force field and the EM-map derived force field, was set to 0.3. Simulations were performed using a 1 fs timestep with nonbonded forces being evaluated every two steps. Born radii were calculated using a cutoff of 14 Å, while the nonbonded forces were smoothed and cut off between 15 and 16 Å. An implicit ion concentration of 0.1 M was assumed with protein and solvent dielectric set to 1 and 80, respectively. A Langevin thermostat with a damping coefficient of 5 ps^{-1} was employed to hold the temperature to 300 K. In the explicit solvent simulation, the ribosome was simulated in a periodic box of TIP3P water [32] including an explicit ion concentration of 0.1 M, with nonbonded forces cut off at 10 Å and long-range electrostatics calculated by PME every four steps. The in vacuo simulation utilized the same parameters as explicit solvent, but without inclusion of solvent or bulk ions, and neither PME nor periodicity were employed.

Each system was minimized for 5000 steps before performing MDFF for 3 ns. For the explicit solvent simulation, an additional 0.5 ns equilibration of water and ions was performed, with protein and nucleic acids restrained, before applying MDFF.

Chapter 4

Cryo-EM Modeling Challenge 2010

4.1 Introduction

Cryo-electron microscopy (cryo-EM) is furnishing images of biological complexes for over 30 years [121]. Continuous progress in instrumentation and methodology has consolidated cryo-EM as a major biomolecular structure determination technique, especially for large macromolecular complexes, the crystallization of which is typically too challenging. Cryo-EM single-particle reconstruction yields now routinely structures at sub-nanometer resolution, in some cases approaching atomic resolution [76, 122, 123, 124, 125, 126, 127, 128]. Even though de novo modeling of atomic structures is now possible in exceptional cases, interpretation of cryo-EM maps typically leverages available crystallographic structures. Approaches to merge structural information of different modalities are collectively called hybrid methods, and are in fact analogous to methods employed for X-ray structure determination. Electron density maps solved by X-ray crystallography at $\sim$3 Å resolution, for instance, clearly do not feature atomic resolution per se, but since the structure of the building blocks (mainly amino acid residues and nucleotides) is known at very high resolution, complete atomic models can be built given such electron density maps.

A number of hybrid methods that combine structural information from X-ray crystallography and cryo-EM have been proposed in the last few years (see Chapter 2 and Trabuco *et al.* [4]), including the molecular dynamics flexible fitting (MDFF) method [3]. Evaluation of the performance of various hybrid approaches in a systematic way is highly desired in the cryo-EM community. Therefore the Cryo-EM Modeling Challenge 2010 was held in which developers of different hybrid methods applied their methods to model the same set of cryo-EM maps and deposited the resultant structures onto a

public-accessible server [34].

As part of the Cryo-EM Modeling Challenge 2010 (`http://ncmi.bcm.edu/challenge`), we have applied MDFF to a number of density maps, namely: GroEL (4 Å), GroEL-GroES complex (7.7 Å), Mm-cpn in its closed state (4.3 Å), lidless Mm-cpn in its open state (8 Å), aquaporin-0 (2.5 Å, electron crystallography), VP6 component of rotavirus (3.8 Å), and a bacterial ribosome (6.4 Å). Below we describe the atomic models obtained with special emphasis on tools to detect and prevent stereochemical errors and application of MDFF to crystallographic data.

4.2 Results

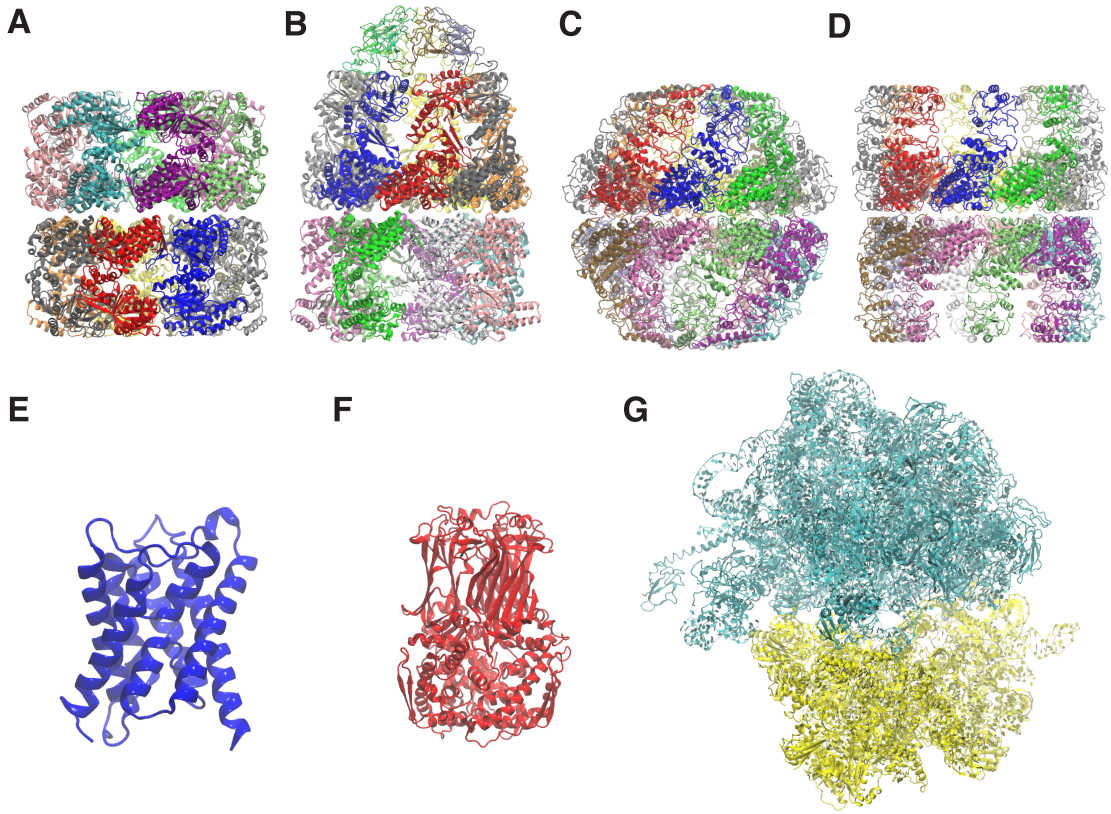

Figure 4.1: Atomic models obtained by applying MDFF to density maps from the Cryo-EM Modeling Challenge 2010: (A) GroEL; (B) GroEL-GroES complex in the ATP-bound state; (C) Mm-cpn in the closed state; (D) lidless Mm-cpn in the open state; (E) aquaporin 0; (F) VP6 component of rotavirus; and (G) a bacterial ribosome.

Atomic models were obtained by applying MDFF to seven density maps provided in the Cryo-EM Modeling Challenge 2010 (Fig. 4.1). In order to assess the quality of the fit for each case,

the cross-correlation coefficient (CCC) between initial/final structures and the corresponding maps was calculated, showing an improvement in all cases (Table 4.1). The classical CCC suffers from an important limitation: all volume within the box encompassing the density map is taken into account in the calculation, such that a significant portion of the data points used does not actually correspond to the macromolecule [3, 4]. In fact, a CCC coefficient close to one can be obtained simply by defining a sufficiently large bounding box. Furthermore, in cases where only a subset of the macromolecular components imaged can be modeled, the unoccupied regions in the map lead to a lower global CCC. Thus, it is usually preferable to use a local CCC measure, in which only regions around the macromolecule are used in the calculation. Initial and final local CCCs presented in Table 4.1 show more clearly the improvement of the fit obtained by employing MDFF simulations.

Table 4.1: Local and global cross-correlation coefficients (CCC) between atomic models and density maps. Local CCCs were calculated using a density threshold of 1 and 0.2 standard deviation above the mean for maps with resolution higher and lower than 5 Å, respectively. Global CCCs are given in parentheses.

System	Resolution	Initial	Final
Aquaporin	2.5 Å	0.58 (0.34)	0.69 (0.39)
GroEL	4 Å	0.32 (0.74)	0.57 (0.86)
GroEL-GroES	7.7 Å	0.44 (0.84)	0.71 (0.90)
Rotavirus	3.8 Å	0.68 (0.56)	0.75 (0.61)
Mm-cpn in closed state	4.3 Å	0.51 (0.74)	0.59 (0.76)
Mm-cpn in open state	8 Å	0.15 (0.39)	0.77 (0.88)
Ribosome	6.4 Å	0.30 (0.75)	0.46 (0.80)

Stereochemical errors

Force fields used in MD simulations do not contain any energy term designed to enforce a given chirality or *cis/trans* peptide bond configuration. In general, both enantiomers or peptide bond isomers can arise, although the two forms are separated by a large energy barrier. Thus, for equilibrium MD simulations starting with error-free atomic structures, stereochemical errors are not expected to arise. In simulations such as MDFF employing external forces, however, it is possible for artifactual chirality or peptide bond configurations to arise, resulting in an atomic model with stereochemical errors. Hybrid methods do apply forces to atoms to drive them into the EM density map and, thus, may introduce stereochemical errors, unless additional measures are taken. To ensure stereochemical integrity of molecular models, software tools were developed to identify, visualize, and interactively correct stereochemical errors in biomolecular structures [129]. These tools are implemented as two

plugins (`Chirality` and `Cispeptide`) to the molecular visualization and analysis package VMD [116]. Given a stereochemically correct starting structure, the plugins can also be used to generate harmonic restraints designed to prevent stereochemical errors from arising. Such stereochemical restraints are included indeed in standard MDFF protocols. Practical details about use of the `Chirality` and `Cispeptide` plugins are covered in Schreiner *et al.* [129], and in a tutorial available on the MDFF website (`http://www.ks.uiuc.edu/Research/mdff`). Since the data of these stereochemical restraints are written out as files for input to the simulation package NAMD [11], they can also be utilized by other MD simulation packages that support user-defined internal-coordinate restraints after adaptation of the data files.

During the Cryo-EM Modeling Challenge 2010, atomic models obtained by different flexible fitting methods were deposited and made publicly available. Analysis of such models can help answer the question of whether hybrid methods not based on MD simulations are also prone to introducing stereochemical errors in the generated atomic models. Table 4.2 presents the number of chirality errors and *cis* peptide bonds identified by the `Chirality` and `Cispeptide` VMD plugins, respectively, for each deposited atomic model obtained with a flexible fitting method, covering a total of five different methods. Structures derived with MDFF or Rosetta [130] were stereochemically correct, whereas the other three investigated methods (DireX [55], Gorgon [125, 131], and Froda [132, 133]) introduced stereochemical errors. The different stereochemical quality of models may stem from the different definition of allowed configurations and the usage of additional restraints in the various methods.

Table 4.2: Chirality errors and *cis* peptides identified in atomic models obtained by various flexible fitting methods, deposited as part of the Cryo-EM Modeling Challenge 2010. The number of chirality errors is given first, followed by the number of *cis* peptide bonds. For rotavirus and ribosome, 3 and 6 *cis* peptides, respectively, involving Pro present in the original structures were not considered. Mm-cpn models obtained by Gorgon contain only a single unit, so the total number stereochemical errors can be obtained by multiplying the given values by 16. Cases for which a model was not deposited are identified with n/a.

System	Resolution	MDFF	DireX	Rosetta	Gorgon	Froda
Aquaporin	2.5 Å	0, 0	0, 0	0, 0	n/a	n/a
GroEL	4 Å	0, 0	0, 4	0, 0	n/a	n/a
GroEL-GroES	7.7 Å	0, 0	194, 60	0, 0	n/a	14, 11
Rotavirus	3.8 Å	0, 0	0, 1	n/a	n/a	n/a
Mm-cpn closed state	4.3 Å	0, 0	0, 0	0, 0	244, 202	n/a
Mm-cpn open state	8 Å	0, 0	n/a	0, 0	267, 257	694, 115
Ribosome	6.4 Å	0, 0	n/a	n/a	n/a	n/a

Apart from peptide bond configuration and chirality, the stereochemical quality of structures can be addressed by comparison with available crystal structure data and structure validation tools like

Molprobity [134]. Here, we chose the 3.8-Å map of the VP6 component of rotavirus to compare with the crystal structure and analyze the comparison with Molprobity. The crystal structure used for comparison is the same as the starting structure used for the fitting (PDB 1QHD [123]). The RMSD between the heavy atoms of the crystal structure and the MDFF model is 1.17 Å, showing close resemblance of the MDFF model to the high-resolution (1.95 Å) crystal structure. Some degree of deviation should be expected due to different molecular environments of the models, i.e., a crystal and a solution environment. Another reason for a significant RMSD value may be a superposition of different conformers present in the cryo-EM map.

An alternative comparison between MDFF model and crystal structure is furnished by backbone dihedrals. The main advantage of using the RMSD between these internal coordinates is that it does not rely on any alignment. Taking all backbone dihedrals into account, the RMSD for the Φ and Ψ angles are both about 21 degrees. Excluding flexible loop regions from the comparison yields RMSDs of 2 and 2.6 degrees for the Φ and Ψ angles, respectively. The reduced RMSD clearly shows that most of the differences stem from the loop regions. Apart from the fact that secondary structure elements were kept restrained during MDFF, a dominant uncertainty in the positions of loop regions should be expected since these regions are less resolved in the EM data and very flexible in MD simulations.

Table 4.3: Comparison of stereochemistry of initial structure, final structure and structure after further minimization without EM data for the map of VP6 component of rotavirus (3.8 Å) as analyzed by the Molprobity web server [134]. Definitions of the stereochemistry indicators are covered in Chen *et al.* [134]

Stereochemistry indicators	Initial	Final	Minimized
Clashscore	0.95	6.01	0.48
Poor rotamers	4.43%	9.49%	1.27%
Ramachandran outliers	0%	3.85%	0.26%
Ramachandran favored	97.44%	87.44%	96.67%
C_β violations	9	101	2
Residues with bad bonds	0%	16.33%	0%
Residues with bad angles	0%	11.48%	0.26%

Although no stereochemical errors were detected in the final fitted structure of the VP6 component of rotavirus, the Molprobity server detected differences relative to the crystal structure. Overall, the MDFF-derived model shows a larger degree of flexibility than observed for the crystal structure (Table 4.3). The origin of the differences is of the same nature as already discussed for the comparison of RMSD in real space. The obtained differences are observed even after a short structure optimization (see Methods for the simulation protocol), suggesting that the variation in structural parameters is compatible with the map and moreover is within a reasonable range as guaranteed by the force field.

A more aggressive optimization using tighter convergence criteria and not taking into account the EM data would be able to reduce even more deviations from the ideal values (Table 4.3). This means, however, that the effect of the temperature, which is present in the EM map, will be eliminated.

Application to electron crystallography

The MDFF method had been developed and validated initially as a real-space refinement method for cryo-EM density maps [3]. Since the Cryo-EM Modeling Challenge 2010 provided a density map obtained by electron crystallography (aquaporin-0 resolved at 2.5 Å), the applicability of MDFF to such data source was tested. The originally solved structure (PDB 3M9I [135]) was thus refined with MDFF. For the purpose of the challenge, which focuses on obtaining structural models based on cryo-EM data, the density map was treated only as an EM map and kept unchanged throughout the refinement process. As a proof-of-principle, a very simple measure was defined to determine if MDFF could potentially provide qualitative improvements. A molecular envelope was defined using a density threshold and, for each residue in the atomic model, the fraction of heavy atoms inside the envelope was calculated before and after MDFF refinement (Fig. 4.2A). Overall, MDFF lead to an increase in the fraction of atoms inside the molecular envelope. Focusing on a few peaks from the plot in Fig. 4.2A, one can visually inspect the effect of MDFF on the local structure at the side chain level (Fig. 4.2B-E). The images show that, in principle, MDFF refinement has the potential to improve the quality of structures solved by electron crystallography.

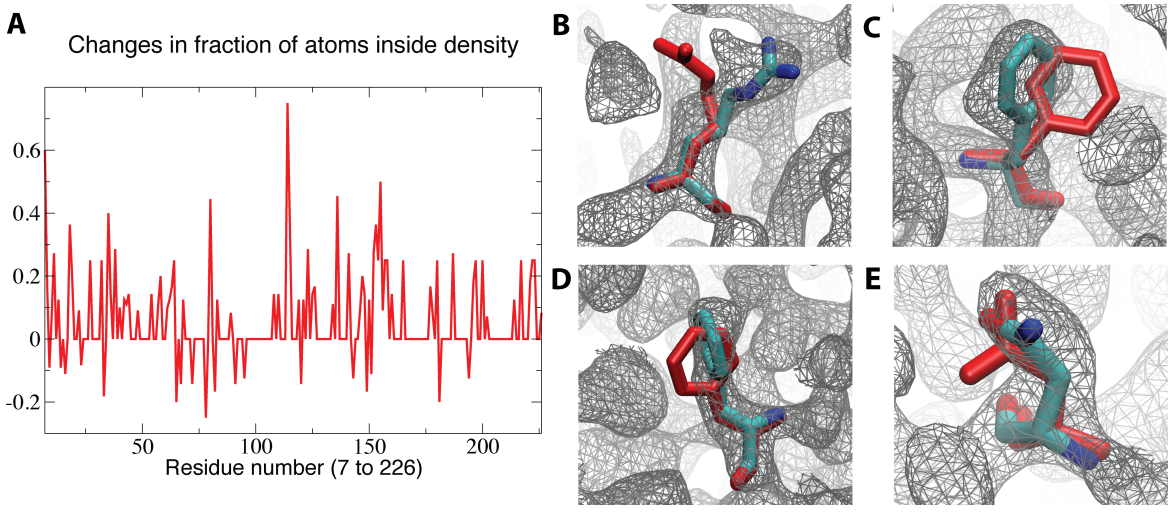

Figure 4.2: MDFF applied to a 3-Å electron crystallography map of aquaporin 0. (A) Per-residue fraction of heavy atoms inside a molecular envelope defined by a density threshold of one standard deviation above the mean. Examples of residues with improved placement of side chains within the density, as identified by peaks in the plot, with initial structure colored in red: (B) Arg-11; (C) Phe-18; (D) Phe-136; and (E) Asn-197.

It should be noted, however, that in crystallographic refinement the phases and, hence, the density map are updated iteratively. Furthermore, one could utilize information like beta factors or even the original diffraction data in the MDFF fitting process. Despite the stated shortcomings, improvement of structure by refinement to a static map showcased here suggests the possibility to develop an extension of MDFF to fit crystallographic data, which is indeed under development [136].

4.3 Discussion

MDFF was applied to obtain atomic models for several density maps determined by the Cryo-EM Modeling Challenge 2010, covering a range of system sizes (single proteins to large complexes) and resolutions (2.5–8 Å). A comparative analysis of atomic models obtained using different flexible fitting methods revealed that certain approaches are prone to introducing stereochemical errors. MDFF protocols include extra harmonic restraints designed to prevent such errors from arising [129].

MDFF benefits from the flexibility of NAMD, its underlying MD package [11], as highlighted by the aforementioned examples. As the MDFF method matures and is applied to a larger range of systems, limitations can often be addressed by introducing special-purpose restraints, such as the ones designed to preserve symmetry [5] or stereochemical correctness [129]. Being part of the very efficient NAMD software, MDFF also benefits from NAMD's high scalability, which is almost linear with system size and number of CPUs. For example, the solvated VP6 component of the rotavirus in the challenge is comprised of 105,537 atoms and the MDFF simulation speed was about 0.66 days per ns using 48 CPUs, while the larger GroEL-GroES system in solvent consists of 594,845 atoms and the speed was about 3.79 days per ns using 48 CPUs, showing almost linear scaling with system sizes. Speed and number of cores required by systems of other sizes should be approximately equal to linear interpolation of these numbers. For instance from linear interpolation of the above numbers one would expect to get about 3.8 days per ns on an eight-core machine for a system of 100,000 atoms. This scalability makes MDFF applicable even to very large systems. In light of ever increasing computational power, computational cost should not be a barrier to the use of MDFF.

Since MDFF is based on MD simulations, any molecule that has been parameterized in MD force fields can be modeled by MDFF. Macromolecules can also be modeled in a realistic environment, namely solvated by water molecules and ions, or even embedded in a lipid membrane [137, 138, 4]. The current release of NAMD 2.8 supports the generalized Born implicit solvent model [6]. Calculations using implicit solvent models reduce the computational cost by avoiding computations associated

with water, which comprises often the largest part of the system. MDFF results obtained with the Born model were shown to provide better agreement with the corresponding results in explicit solvent than in-vacuo fittings [6]. Since the implicit solvent feature was not yet available at the time of the challenge, the benefits cannot be showcased here. Readers are referred to Chapter 3 and Tanner *et al.* [6] for more information on implicit-solvent MDFF.

The Cryo-EM Modeling Challenge 2010 was the first community effort to evaluate the performance of different hybrid methods in a systematic way. Structural biologists wishing to apply hybrid methods now have more information to decide which approach is most appropriate to their particular problem. Although concrete conclusions are not made in this first attempt of the challenge, it is expected that the next challenge will provide more information for potential users of hybrid methods. Such efforts also motivate researchers to push the boundaries of their methods by applying them to new kinds of problems. As an example, the first proof-of-principle application of MDFF to crystallographic data resulted from the challenge. The hybrid modeling field as a whole stands to benefit greatly from such community efforts. Method developers can learn from each other and further improve their own methods in an iterative way, and opportunities arise for combining multiple methods to tackle modeling tasks of ever increasing difficulty.

4.4 Methods

The following structures were used as the starting point for the MDFF simulations presented in this paper: PDB 3E76 [139] was fitted into the 4-Å GroEL map [122]; PDB 2C7D [71] was fitted into the 7.7-Å GroEL-GroES map [71]; PDB 3LOS [76] was fitted into the 4.3-Å map of Mm-cpn in a closed state [76]; a homology model of a lidless Mm-cpn complex was built with Modeller [140] using PDB 3LOS as a template and fitted into the 8-Å map of lidless Mm-cpn in an open state [76]; PDB 3M9I [135] was fitted into the 2.5-Å aquaporin 0 map [135]; PDB 1QHD [141] was fitted into the 3.8-Å map of a VP6 component of rotavirus [123]; and PDB 2WDG/2WDI [142] was fitted into the 6.4-Å ribosome map [143] after removing bound factors. Each structure was first rigid-body docked into the corresponding density maps using `colores` from the Situs package [51].

Aquaporin-0 was embedded in a POPE lipid bilayer of size $100\,\text{Å} \times 100\,\text{Å}$, with TIP3P [32] water molecules added with a 15-Å padding in the direction orthogonal to the membrane. The ribosome system was simulated in vacuo, whereas the remaining structures were solvated in a TIP3P water box with a 10-Å padding in all directions. All systems simulated in explicit solvent had their total charge

neutralized by adding Na^+ or Cl^- ions.

MD simulations were performed with a development version of NAMD 2.8 [11]. The CHARMM27 force field with CMAP corrections [79, 80] was used for all systems except for the ribosome, which was simulated with the AMBER99 force field [117] including the SB [118] and BSC0 [119] corrections, and converted to CHARMM format [16] to allow for system building with VMD [116]. The dielectric constant was set to 80 for the ribosome simulation in vacuo and to 1 for the remaining systems. Temperature was maintained at 300 K using Langevin dynamics with a damping constant of 5 ps^{-1}. For the aquaporin system (embedded in a membrane), pressure was maintained at 1 atm employing a Nosé-Hoover-Langevin piston with a decay period of 200 fs and time constant of 100 fs. The RESPA [81, 82] multiple-time-stepping algorithm was used with an integration time step of 1 fs, short-range forces calculated every 2 fs, and long-range electrostatics calculated every 4 fs. Nonbonded interactions were calculated with a 10-Å cut-off. Long-range electrostatic forces were computed by the particle-mesh Ewald summation method using a grid spacing smaller than 1 Å. All systems simulated in explicit solvent were subjected to restrained MD simulations prior to flexible fitting, with harmonic restraints first applied to all protein atoms, followed by restraints applied only to backbone atoms, thus allowing water, ions, lipids (in case of aquaporin), and side chains to equilibrate.

The grid scaling, an MDFF parameter that controls the balance between the map-derived potential energy term and the normal MD force field, was set to 0.3 [3]. Harmonic restraints were applied to enforce correct chirality and peptide bond configuration [129]. Each MDFF simulation was performed until convergence of the protein RMSD, followed by 3,000 steps of energy minimization in the presence of the density-derived MDFF energy term and a grid scaling of 10. For symmetry-restrained MDFF simulations [5], only C_α atoms experienced symmetry restraints, with the force constant linearly increased from 0 to 10 kcal/mol/$Å^2$ throughout the simulations.

Chapter 5

Mechanism of tetracycline resistance by ribosomal protection protein Tet(O)

5.1 Introduction

Tetracycline (Tc) is a broad-spectrum antibiotic active against both gram-positive and gram-negative bacteria, and used in a variety of medical and veterinary applications [35]. It targets the small subunit of the bacterial ribosome [144], with the primary high-affinity binding site located beneath helix 34 of the 16S rRNA, in a crevice between head and platform of the 30S subunit, in close vicinity of the A-site codon [144]. In this binding site, Tc partially occupies the space where an aminoacyl-tRNA would approach the A-site codon during the decoding process. Thus, Tc binding prevents any entering aminoacyl-tRNA from being recognized by the codon in the messenger RNA and thereby incapacitates the ribosome for protein synthesis [145].

Resistance to Tc is mediated through one of several mechanisms: Tc efflux, protection of the Tc binding site by binding of specific cytoplasmic proteins to the ribosome, Tc modification, or modification of 16S rRNA at the Tc-binding site [146]. These mechanisms are facilitated by over 20 different groups of tetracycline-resistance proteins [147]. Several of these proteins - the best-known being Tet(M) and Tet(O) - are paralogs of the translational GTPase EF-G [148] and actively remove Tc from the ribosome in a GTP-hydrolysis-dependent fashion [36, 37].

Detailed information on the Tet(O)-induced conformational changes of the ribosome to disrupt Tc binding is essential for understanding their mechanism of action. A previous Tet(O) cryo-EM study [149], with a density map at a resolution of 16 Å, allowed the visualization of Tet(O) bound to the *Escherichia coli* ribosome, revealing that Tet(O) indeed has a similar shape as EF-G and

binds to the same general site within the intersubunit cavity of the ribosome. Overall, the ribosomal complex was seen to closely resemble the conformation of the EF-G-70S ribosome complex stabilized with GDPNP. Sequence homology shows that Tet(O) shares the five structural domains with EF-G, the first two containing the GTP-binding site and being close to the GTP-associated center of the 50S ribosomal subunit, while the other three domains are distinct for Tet(O) and associated with its specific functions [149, 150].

The high degree of sequence homology shared by Tet(O) and EF-G [151] makes it possible to build an atomic model of Tet(O) based on the X-ray structure of EF-G, using a cryo-EM map as constraint, provided its resolution is sufficient. The existing 16-Å reconstruction of Tet(O)-70S [149] is unsuitable for this purpose. In this study, we obtained an improved cryo-EM map of the *E. coli* 70S ribosome in complex with GDPNP-bound Tet(O) from *Campylobacter jejuni*, the best characterized ribosomal protection protein, with a resolution of 9.6 Å. Guided by this higher-resolution map, a map-fitted atomic model of the ribosome-Tet(O) complex has allowed us to determine the binding sites between Tet(O) and the ribosome. Our results indicate that the critical residues of Tet(O) would clash with Tc if both were present in the same ribosome complex; thus the presence of Tet(O) is poised to disrupt the binding of Tc. These critical residues are located in three loops of domain 4, whose positions with respect to the ribosome are different than for EF-G [152]. Moreover, our structural results are strongly corroborated by our mutational and biochemical data. Multiple tests of the three Tet(O)'s loops with either a single Ala mutation in one of the three loops, or a replacement of any of the three residues tipping the three loops by a glycine, resulted in loss of Tet(O) functionality as measured by minimum inhibitory concentration (MIC) of the antibiotic resistance of *E. coli* strains. Together, these results allow us to understand the structural basis for the Tc-resistance mechanism on the molecular level.

5.2 Results

Cryo-EM density map of Tet(O) bound with the 70S ribosome

We obtained a three-dimensional cryo-EM density map for the complex of Tet(O) bound with the 70S ribosome in the presence of the nonhydrolyzable GTP analog GDPNP at a resolution of 9.6 Å. The density for Tet(O) in this improved map is at a position which agrees with that described in the earlier study [149], but boundary and shape are better defined in the context with the 70S ribosome (Fig. 5.1).

To date, Tet(O)'s structure has not been solved by crystallography. In the present study, the

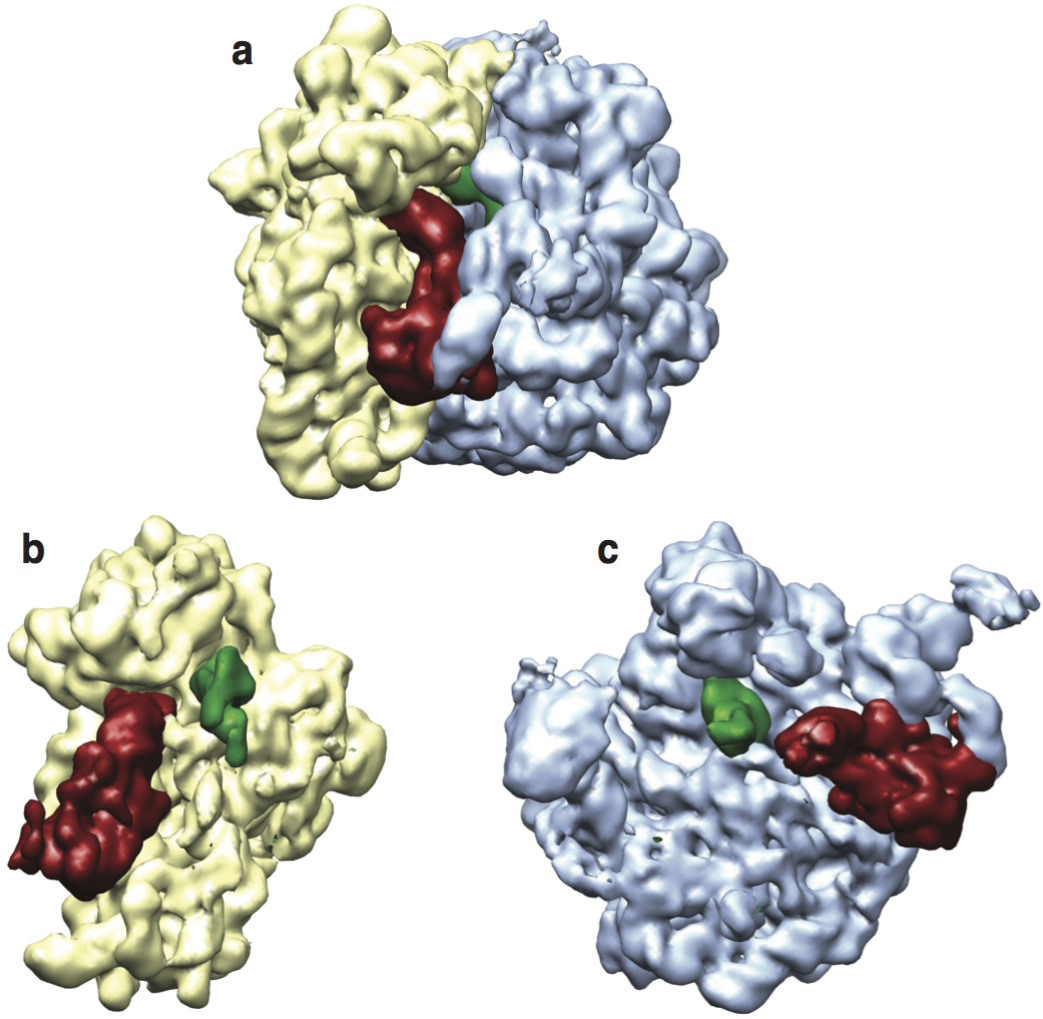

Figure 5.1: Cryo-EM maps of the 70S ribosome from *E. coli* bound with Tet(O). (a) 70STet(O)GDPNPfMet-tRNA complex. The map is segmented to show the 30S subunit (yellow), the 50S subunit (blue), Tet(O) (red) and the P-site tRNA (green). The same color scheme is used for panels (a)-(c). (b) Density for the 30S subunit with Tet(O) and P-site tRNA. (c) Density for the 50S subunit with Tet(O) and P-site tRNA.

improved resolution of the density map allowed us to build an atomic model of Tet(O), guided by its homology to EF-G, and characterize its binding interactions with the ribosome. First, a sequence alignment of *C. jejuni* Tet(O) and *Thermus thermophilus* EF-G was created, guided by the crystal structure of EF-G (PDB accession code: 2WRI) using the 3D-coffee software [153]. To compare site-specific sequence conservation across the entire Tet(O) and EF-G families, a dataset of sequences belonging to both of these families was assembled. Tet(O) sequence homologs, representing the Tet family of translational GTPases (trGTPases), were retrieved from the NCBI using BlastP. These sequences were aligned with a dataset of previously identified sequences from the EF-G family [154],

and consensus sequences were generated for each family. The sequence alignment results indicate a high sequence similarity (50%) between Tet(O) and EF-G, a firm basis for using homology modeling. In all these sequences, the highly conserved nucleotide-binding motifs are perfectly aligned, and the sequence of Tet(O) is divided into five domains, corresponding to the domains defined by the sequence of EF-G. The alignment of *C. jejuni* Tet(O) and *T. thermophilus* EF-G was used to create a homology model of Tet(O) using the program Modeller [155].

Our model of Tet(O) closely resembles the structural features of EF-G which is in a 70S ribosome-bound GTP form (Fig. 5.2). The Tet(O) model together with the X-ray structure of the 70S ribosome [156] was fitted into the entire cryo-EM map for the complex using the Molecular Dynamics Flexible Fitting method [4]. The fitted structure, as shown in Fig. 5.3, closely captures the conformation of the entire complex as formed in the density map, which allows us to characterize the interactions between Tet(O) and the ribosome in detail. We validated the fitting structure by using a different crystal structure of the 70S ribosome as the starting structure for the MDFF (2WRI, 2WIJ). The resultant structure of Tet(O) as well as the bases re-arrangement in the 16S rRNA are in good agreement with the model presented here. The RMSD of the two 70S ribosomal structures is just 1.5 Å, and just 1.7 Å for all atoms of Tet(O).

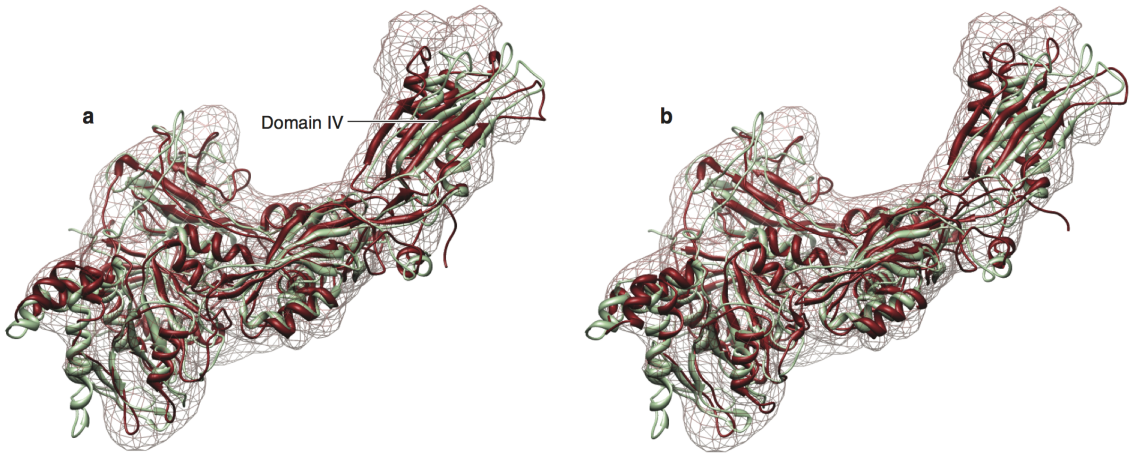

Figure 5.2: Homology model of Tet(O). (a) The initial Tet(O) model (red) and the EF-G structure ((green, PDB code: 2WRI) are shown as overlaid ribbons in the segmented density map (red mesh). The three loops in domain IV are seen to be positioned differently in the two structures. (b) The map-fitted structure of Tet(O) (red) is superimposed with the crystal structure of EF-G, showing the adjustment in the orientations of the three loops of Tet(O) into the density through the map fitting process.

Interactions between Tet(O) and the ribosomal 30S subunit

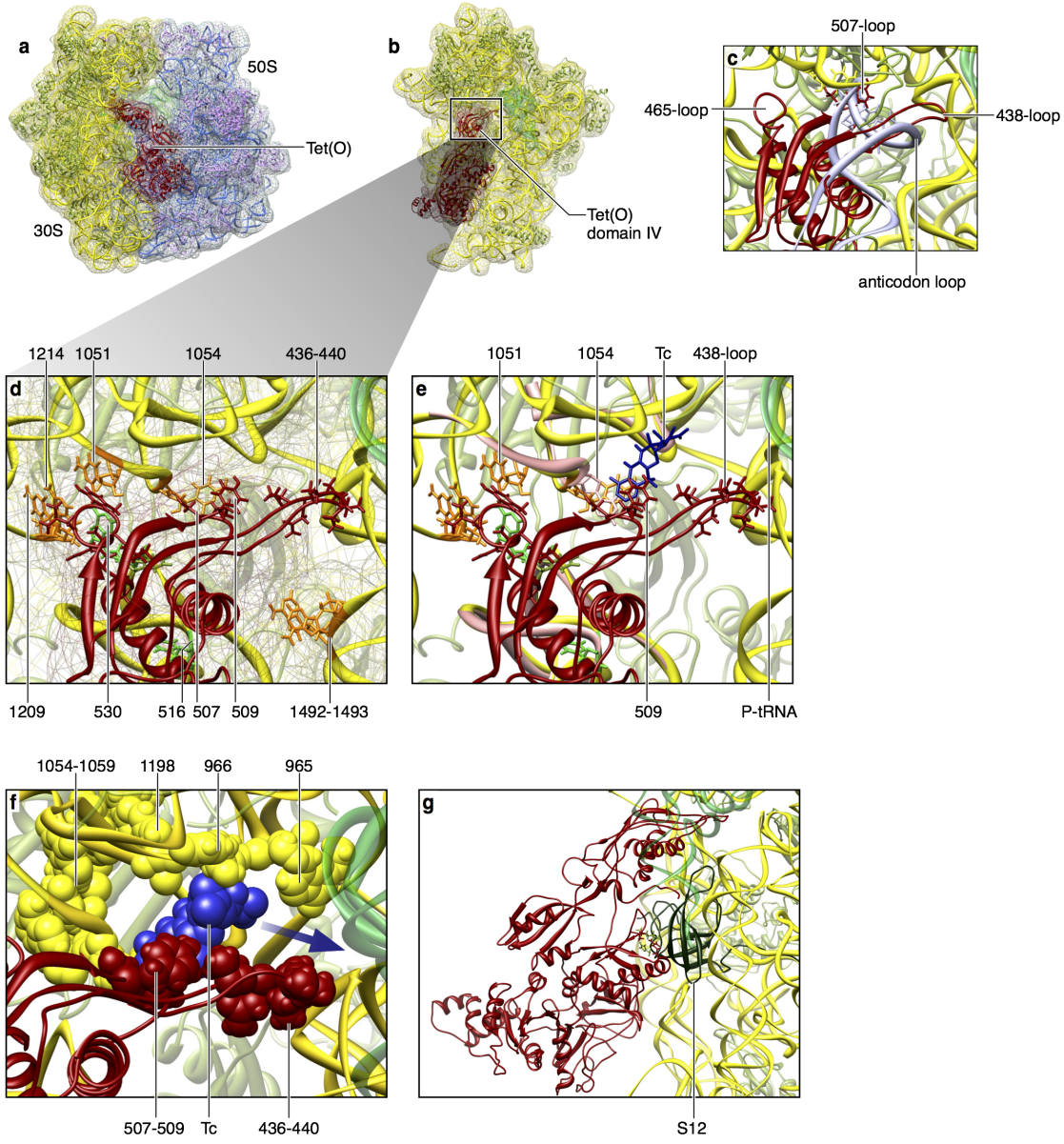

Figure 5.3: Cryo-EM map-fitted atomic model of the 70S ribosome bound with Tet(O). (a) Cryo-EM map, displayed as transparent mesh, with atomic model displayed as ribbons. The 16S rRNA is in yellow, proteins in the 30S subunit in green, 5S and 23S rRNA in blue, and proteins in the 50S subunit in pink. Tet(O) is in red, P-site tRNA in dark green. (b) The 30S subunit with Tet(O) and P-site tRNA shown in (a) is displayed after a clockwise rotation around vertical (in plane) axis by 90°, to show the interface with the 50S subunit. (c) Tet(O)'s domain IV (red), superimposed with the anticodon stem and loop of an A/T-site tRNA (grey) in the decoding region of the 30S subunit. The overlap between Tet(O) and the tRNA occurs at the 507- and 438-loops. (d) Zoomed-in view of the area marked by green box in (b). (e) Portion of the 16S rRNA (pink) surrounding Tc, superimposed on the 16S rRNA (yellow) in the current Tet(O)-bound model. The 507-loop is seen to clash with the site for Tc (blue). Nucleotides 1051 and 1054 in the 16S rRNA are seen to be re-oriented in the two structures. (f) The 507-loop, 438-loop and a portion of the 30S subunit including nucleotides 966, 1196 and their surrounding nts form a structural corridor for a possible release of Tc. (g) Tet(O)-S12 contact sites. S12 is shown in dark green.

Tet(O) binds with the ribosome on both the 30S and 50S subunits. On the 30S subunit, the three loops in domain IV of Tet(O), namely the 465- , 507- and the 438-loop, insert into the head-platform crevice, spanning across about 30 Å (Fig. 5.3c-d). The large spatial expansion of the three loops may allow them to play various roles in their binding with the ribosome, in jointly conveying Tc resistance. The 507-loop and the 438-loop occupy approximately the position where the anticodon stem-loop of an A-site tRNA would occur in a normal translating ribosome, while the 465-loop extends into the structural pocket formed between nucleotides 1055 and 1209 within helix 34 (Fig. 5.3c). Although this third loop is placed outside the decoding center at the A site, it apparently plays a role in restraining the structural flexibility between the 1055 and 1029 ends.

The 507-loop, which is located in the middle of the lined-up three loops, appears to have the most direct, and most crucial role in bestowing tetracycline resistance as evident from its position relative to Tc. The position of Tc in the 30S subunit, as revealed by the crystallographic study [144], was mapped to the current structure of the 30S subunit (Fig. 5.3e). This position would result in a spatial clash between residues 507-509 and Tc if both were present in the same complex, i.e., either Tc or the Tet(O)'s 507-loop would exclusively occupy the same space. Unlike the 507-loop, the 438-loop is not involved in the interactions with Tc, located about 15 Å away from Tc, but residue 437 is located in the near vicinity of nucleotides 1492-1493 in the 16S rRNA, where the decoding interaction network is observed in a normal translation ribosome [157]. We observe that this loop and nucleotides around 966 and 1196 of the 16S rRNA with their surrounding nucleotides form a corridor, which starts at the position of Tc and leads to the outside of the ribosome (Fig. 5.3f). The dissociating Tc molecule has to navigate this corridor, which acts as the only pathway for the molecule's release from the ribosome, which might explain the high activation energy of the process [37].

The presence of the 465-loop inside the structural pocket near nucleotides 1051 and 1209 in the 16S rRNA leads to a local distortion of the 16S rRNA, in the immediate vicinity of the Tc-binding site (Fig. 5.3d). According to the Tc-bound 30S subunit, this loop occupies the position of nt 1209. The presence of the 465-loop of Tet(O) at the base of the 30S subunit's beak prevents the head of the 30S subunit from rotating, a motion required in the normal course of mRNA-tRNA translocation [149].

Tet(O) interacts with the 30S subunit protein S12 closely at domain III, and possibly at domain IV of Tet(O) (Fig. 5.3g). Residues 358 and 379 in domain III of TetO seem to be directly to interact with S12's residues 74-76. The S12 residues 483 and 517 are located at the base of the 507- and 465-loops of Tet(O), respectively. This interaction between S12 and Tet(O) at multiple sites is similar to the bridge-like connection that S12 forms with two sites of the A/T-site tRNA in the EF-Tu-tRNA

complex, as previously described for the X-ray structure of the ternary complex-bound ribosome complex [158].

It is worth noting that the Tet(O) 438-loop is located close to the mRNA around the A-site codon, as shown in our map-fitted structure. The point of closest approach (within 3 Å) occurs between residue 438 and the second nucleotide for the A-site codon.

Functional importance of Tet(O) domain 4 loops

We validated our structural results by testing the effects of mutations of Tc-interacting regions on Tet(O) functionality *in vivo*, as judged by measuring the antibiotic resistance of *E. coli* strains transformed with plasmid-expressed Tet(O) mutants. These mutations and deletions were targeted in the three characteristic loops in domain 4 of Tet(O), at those residues that are either directly or closely involved in the interactions in the Tc-binding site. The effects of the mutations and deletions on Tet(O) functionality were measured through the reduction in the minimum inhibitory concentration (MIC) in each case.

Table 5.1: Inhibition of Tet(O) activity in mediating resistance by different mutations.

Loop	Mutation	Inhibition of Tet(O) activity (in %)
507-loop	Y507A	83
	YSP507-509G	100
465-loop	L466A	33
	S472A	67
	LGY466-468G	100
438-loop	P438A	83
	VPP436-438G	100

The results of the mutation experiments are presented in Table 5.1. All mutations in the three characteristic loops, namely the 507-loop, the 465-loop and the 438-loop, resulted in significant inhibition of Tet(O)-mediated resistance. These results demonstrate that the integrity of the residues identified in our cryo-EM reconstruction as binding sites is crucial for Tet(O) functionality. The mutation results are fully consistent with our cryo-EM structural analysis.

Interactions between Tet(O) and the ribosomal 50S subunit

On the 50S subunit side, Tet(O) contacts the GTPase-associated center, between residues 619-620 in domain III and nucleotides 1066 in helix 43 of the 23S rRNA. The residues of Tet(O) responsible for GTP hydrolysis form a similar structural pocket as in EF-G, which surrounds the sarcin-ricin loop

(nts in the 2662-loop). It is in this pocket where the hydrolysis takes place. In the Tet(O) complex, the closest distance is found to be between nucleotide 2663 and residue 40, about 4 Å. The current map-fitted structure would need to be slightly adjusted around the GDPNP-binding position if GDPNP was included in the fitted structure. Interestingly, helix 69 of the 23S rRNA, which was repeatedly found to be involved in the interactions with other ribosomal factors such as the EF-Tu-aminoacyl-tRNA complex and EF-G [152, 158], is positioned beyond a bonding distance from Tet(O). The closest point (nucleotide 1914) is about 6 Å away from Tet(O). Therefore, in the current Tet(O)-ribosome complex, helix 69 has no direct involvement.

Tet(O) binding-induced conformational changes in 30S

The overall ribosome conformation displays no intersubunit rotation when Tet(O) is stalled on the ribosome in the presence of GDPNP. Essentially, binding with GDPNP allows Tet(O) to remain in its GTP form. This ribosome conformation agrees with the X-ray structure of the 70S ribosome bound with EF-G-GDP-fusidic acid [152]. However, a notable change of ribosomal conformation occurs in a rearrangement that was identified using chemical footprinting [159]. In the complex with Tet(O), the backbone shape of helix 34 deviates from the normal structure as formed in the X-ray structures of the 30S subunit either bound or unbound with Tc (Fig. 5.3e). This RNA fragment in helix 34 includes a single-nucleotide bulge at 1051, as well as two unpaired nucleotides 1054-1055, and connects the rest of helix 34 on the two ends. These unpaired nucleotides naturally provide structural flexibility, which is evidently exploited in the binding of Tet(O) to the ribosome. Our structure shows that residue 507 in domain IV and nucleotide 1054 in helix 34 would spatially clash if the fragment of helix 34 remained in its normal position. Therefore, the binding of Tet(O) clearly causes a change in that region of backbone.

Following the positional change in nucleotide 1209, the backbone shape of the fragment around nt 1051 also must adjust. The base of nucleotide 1054 in particular seems to play a crucial role in holding tetracycline on the ribosome based on the crystal structure by Ramakrishnan and coworkers (PDB ID 1HNW, [152]). Tetracycline adheres to the 30S subunit through multiple hydrogen bonds with nucleotide 1054. The position of nt 1054 in the current structure is re-oriented from being in the Tc-bound ribosome, otherwise it would clash with the 507-loop in Tet(O).

Another Tet(O) binding-induced change in the ribosome occurs at helix 18, including nucleotide 530, which is crucially involved in the network of bases during the normal decoding process. The Tet(O) residue 512-513 in the map-fitted structure are within bonding distance from nucleotides

516-519 of helix 18. This clash is resolved by the re-orientation of nucleotide 516, which appears in a twisted conformation of helix 18 in our map-fitted structure. In this unusual conformation of helix 18, its nucleotide 530 is also reoriented. Evidently, the original orientation must be restored if translation on the ribosome were to resume after the Tet(O) is released from the ribosome along with Tc. The correction would have to be spontaneous after Tet(O) release. This subsequent untwisting process, which we must assume as part of the bacterial rescue of the ribosome, indicates a substantial flexibility of the structure of helix 18. In addition, the bases of nts 1492-1493 in the 16S rRNA flip out toward the Tet(O)'s 438-loop (Fig. 5.3); the flipped-out conformation that was also observed when a codon-anticodon recognition takes place in a translating ribosome [157].

5.3 Discussion

In this study, the structure of the 70S-Tet(O) complex in the presence of GDPNP has allowed us to visualize the details of binding between the 70S ribosome and Tet(O). In the GDPNP-bound form Tet(O) was earlier shown biochemically to promote release of Tc from the 70S ribosome [159, 160], and the elucidation of the current structure therefore provides direct functional insights into the mechanisms of Tet(O)-mediated Tc resistance.

On the 30S subunit side, Tet(O) is positioned close to the site where Tc has been found in the X-ray structure. The 30S subunit-Tet(O) contact sites we have identified in the present study can be divided into two categories: (1) those which lead to a clash with the space for the binding of Tc via the 507-loop and (2) those which disrupt the structure of the 16S rRNA around the Tc-binding site via the 465-loop toward nucleotide 1209 in 16S rRNA. As in EF-G, the GTP-binding site in Tet(O) is located at the GTPase-associated center. These contacts between the Tet(O) and the ribosome seem to collectively play the role of preventing or reversing the binding of Tc to the ribosome.

The most direct effect of Tet(O) binding in preventing Tc from binding to the 30S subunit seems to be the result of a competition between residues 507-509 of Tet(O) and Tc for the same space. When Tet(O) enters into the Tc-bound ribosome, the 507-loop cannot be settled into the ribosome complex because Tc already occupies the close vicinity of nucleotide 1054 and forms multiple hydrogen bonds with the ribosome. The competition for the same site guarantees that Tet(O) and Tc cannot coexist in the ribosomal complex. In addition, binding of Tet(O) disrupts the ribosome structure and reshapes the geometry of the backbone where Tc is anchored (Fig. 5.3e). With this disrupted backbone structure, the nucleotides involved in binding with Tc are reoriented, and thus, Tc loses its

bonds with the ribosome. Interestingly, the binding of Tc does not change the backbone shape from its shape in the Tc-free ribosome.

The question arises is by which molecular mechanism Tc inhibits normal translation in the ribosome. One may ask why a Tc-bound ribosome does not accept an entering EF-Tu-bound aminoacyl-tRNA complex, but does accept Tet(O) even though the EF-Tu-aminoacyl-tRNA complex forms a shape highly similar to that of Tet(O). Our current study provides some insights to answer this question. If an aminoacyl-tRNA bound with EF-Tu enters into the ribosome, its anticodon loop must reach the codon site. In the presence of Tc, a primary portion of the space for the anticodon loop is already occupied by Tc (Fig. 5.3e) which causes a decisive rejection of the aminoacyl-tRNA from the ribosome before codon-anticodon recognition can take place. In contrast, Tet(O) enters the ribosome with less demand for space in that region than the anticodon loop of the tRNA; the available space provides an opportunity for Tet(O) to be admitted to the factor binding site, important for subsequence GTP hydrolysis.

Tet(O), a GTPase, possesses a structure very similar to that of ribosomal GTPase, elongation factor G. The structural similarity suggests an analogy of their GTP-hydrolysis-induced conformational changes which enable the two ribosomal proteins to perform their respective biological functions. The structural effects of the EF-G-associated GTP hydrolysis on the ribosome have been extensively studied [161, 19, 162]. It is known that GTP hydrolysis induces substantial conformational changes in the ribosome. The ribosome's effect on the conformation of EF-G is substantial, as well, causing domain IV to be reoriented relative to the other domains, as shown by cryo-EM [19] and X-ray structures [152] of the ribosome bound with EF-G in the presence of fusidic acid. The antibiotic fusidic acid traps EF-G in a conformational intermediate between the GTP and GDP forms. In this translocational complex, a contact observed between the 507-loop of EF-G and the P-site tRNA seems to be essential for the translocation of tRNA based on the significant conformational flexibility of this loop [152]. This flexibility allows the loop to participate in the major dynamic motion of the entire domain IV as the GTP hydrolysis takes place. Thus, we see this loop as a functionally required structural element in EF-G. The structural similarity of Tet(O) to EF-G suggests structural flexibility in the equivalently positioned loops of Tet(O). Accordingly, we predict that GTP hydrolysis in Tet(O) results in extensive conformational changes in the distal loops of domain IV, particularly in the three flexible loops. By combining structural and mutational analyses, the present study provides structural insights into how the three loops in domain IV (see Fig. 5.3) might cooperate to expel Tc from the ribosome: the 465-loop is responsible for distorting the backbone shape at nucleotides 1051-1054 of

16S rRNA, which weakens or abolishes the binding of Tc at this site with the RNA; the 507-loop in the middle of these three directly pushes Tc out of the ribosome; and the 438-loop along with nucleotide 966 and 1196 should form a corridor allowing Tc to exit.

After completion of this work, a cryo-EM reconstruction of a 70S-Tet(M) complex was published [163]. Our results agree with the results by Wilson and coworkers in all essential details, as expected based on the high sequence homology between Tet(O) and Tet(M).

5.4 Methods

For assessing similarity between and among the EF-G and Tet(O) families of GTPases, 313 Tet(O) sequences were retrieved from the NCBI RefSeq database using BlastP with *C. jejuni* Tet(O) as the query. 171 EF-G sequences were taken from the data set of [154] and EF-G and TetO sequences were aligned with Mafft [164]. Consensus sequences were calculated with Consensus Finder [154]. To visualize a subset of aligned representative sequences from across the diversity of the Tet family, phylogenetic analysis was carried out using FastTree [165].

For homology modeling, an alignment of Tet(O) with EF-G was made with 3Dcoffee [153], taking into account the structure of *Thermus thermophilis* EF-G (PDB ID 2WRI [166]), for the placement of insertions and deletions. The sequence alignment result was used to build an atomic model of Tet(O) using Modeller [155]. This atomic model and the X-ray structure of a 70S ribosome including a P-site tRNA (PDB codes: 2J00, 2J01 [156]) were together fitted into the cryo-EM map by means of MDFF [4], assuming a generalized Born implicit solvent as implemented in NAMD [6].

Experimental details of the cryo-EM and the mutation experiments are described in [8].

Chapter 6

Dynamic behavior of trigger factor on the ribosome

6.1 Introduction

During translation the nascent polypeptide chain emerges from the ribosomal exit tunnel and is released into the crowded environment of the cytosol where processes such as protein folding and targeting for translocation must occur. These processes are assisted by protein factors such as the signal recognition particle (SRP) and its receptor (SR) in the case of targeting, and different chaperones and chaperonins in case of protein folding. Whereas in eukaryotes a whole repertoire of chaperones (such as RAC, NAC, SSB) interacts with the ribosome near the tunnel exit on the large ribosomal subunit, in eubacteria trigger factor (TF) is the only chaperone, which directly interacts with the ribosome. TF is composed of three domains: an N-terminal ribosome binding domain (RBD), the peptidyl-prolyl cis/trans-isomerase (PPIase) domain or head domain, and the C-terminal domain located between the RBD and PPIase domain in the TF structure and forming the arms 1 and 2 [167]. The RBD and the C-terminal domain harbor the main chaperone activity of TF, whereas the PPIase domain is dispensable for the chaperone activity *in vivo* [168, 169]. Nevertheless, more recent reports suggest that the PPIase domain provides a second chaperone interaction side for the nascent chain at least for some proteins [170, 171]. Most of our knowledge about the involvement of chaperones in protein folding originates from *in vitro* studies, where the refolding of chemically denatured substrate proteins was monitored in the presence of chaperones with various techniques. During recent years though, many studies have shown that in eukaryotes, as well as in prokaryotes, protein folding starts already on the ribosome, while the C-terminal part of the polypeptide still gets extended by the peptidyl-transferase center. Furthermore, several recent studies have shown that the formation of secondary structures, such as α-helices, already begins within the ribosomal exit tunnel [172, 173, 174, 175].

To date, three crystal structures of TF ribosome binding domains (RBD) in complex with 50S ribosomal subunits (one heterologous pair of *E. coli* TF and archaebacterial 50 S from *Haloarcula*

marismortui and two homologous pairs of TF and 50 S from *Deinococcus radiodurans*) have revealed the overall positioning of the RBD on empty ribosomes [176, 177, 167]. Docking of the crystal structure of a complete TF molecule onto the structure of the 50S associated RBD indicated that the chaperone is localized in a position so that it "arches" over the ribosomal tunnel exit [167]. A cryo-EM structure of *E.coli* ribosomes carrying a stalled nascent chain of a non-physiological substrate (SH3, which can fold independently of TF) with a TF molecule fixed to the nascent chain via a covalent disulfide bond in close proximity to the tunnel exit, confirmed the localization of the chaperone over the ribosomal tunnel exit. Nevertheless, due to the limited resolution of this structure (19 Å), the interaction of TF with the nascent chain could not be studied in detail yet [178].

Therefore, we aimed at the visualization by cryo-EM of ribosome-bound TF when interacting with a physiological substrate emerging from the ribosomal tunnel. For this purpose we choose the physiological substrate galactitol-1-phosphate dehydrogenase (GatD) as nascent polypeptide chain that is known to interact with TF and exhibits two hydrophobic stretches [171, 179].

We show, that TF interacts with ribosome nascent chain complexes (RNCs) carrying this substrate with high affinity. The cryo-EM structure exhibits a high resolution for the ribosome and the N-terminal RBD of trigger factor, whereas the middle domain and the head (PPIase) domain of TF are less-well resolved or even non-detectable due to increasing degrees of flexibility. The resolution of our RNC-TF-NBD map (7.7 Å) enabled us to visualize the nascent chain within the ribosomal tunnel and to even trace it onto the surface of the TF-RBD. We show that the two alpha helices of the RBD undergo a conformational change when bound to the translating ribosome. This rearrangement causes the presentation of a new hydrophobic surface on the NBD, which serves as interaction site for the hydrophobic stretch within the emerging nascent chain. Using microscale thermophoresis we show that the affinity of TF for RNCs increases when the nascent chain is longer that may indicate an altered conformation of TF. We confirm this finding with a second cryo-EM structure, where TF in complex with RNCs carrying a longer nascent chain of GatD gained rigidity partially due to additional interactions of TF with either the nascent chain or the ribosomal surface. Molecular dynamics simulations reveal the molecular interactions between TF and the nascent chain, elucidating how the length of the nascent chain modulates the conformation of RNC-bound TF and the binding/unbinding of TF with RNCs.

6.2 Results

Cryo-EM structures of TF-RNC complex

Flexibility of RNC-bound TF inferred from cryo-EM data. We generated RNCs carrying a GatD nascent polypeptide stalled by the TnaC stalling sequence *in vivo* and subjected these complexes to cryo-EM and single particle analysis. A cryo-EM map of an overall resolution of 7.7 Å was obtained. To our surprise this density did not account for the entire TF molecule but rather represented only parts of the RBD of TF. When we low pass filtered the density map to a resolution of 16.5 Å the C-terminal domain with its arms became visible (Fig. 6.1b), whereas some densities of the PPIase domain show up when we further low pass filtered the density map to a resolution of 35 Å (Fig. 6.1c), indicating a gradual increase in flexibility within TF from the RBD over the C-terminal domain towards the PPIase domain. The crystal structure of full length TF exhibits two possible pivot points that are likely to explain the observed flexibility. The RBD is connected via an extended unstructured loop (V111 to T133) with the PPIase domain. This loop is packed against another loop connecting the long α-helix forming the back with arm1 (D299 to P303). The second pivot point could be represented by the unstructured loops (A149 to K154 and E241 to L248) connecting the PPIase domain with the remaining of the TF molecule. These pivot points are in agreement with our MD simulations (see MD section).

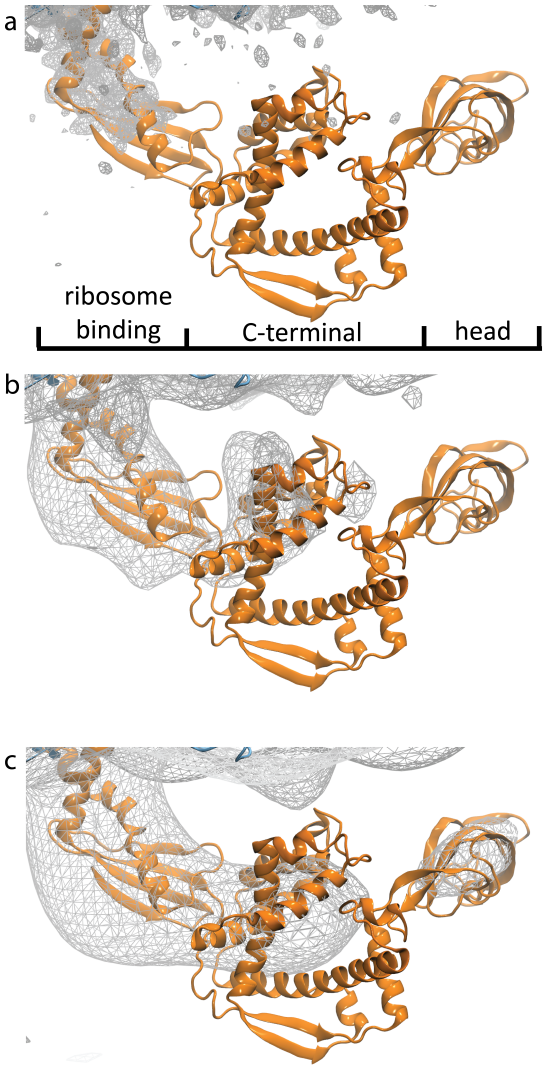

Figure 6.1: MDFF model of the ribosome-bound trigger factor. Densities of the EM maps for the GatD-RNC[85] complex at 7.7 Å, 16.5 Å and 35 Å resolution are shown in (a), (b) and (c), respectively.

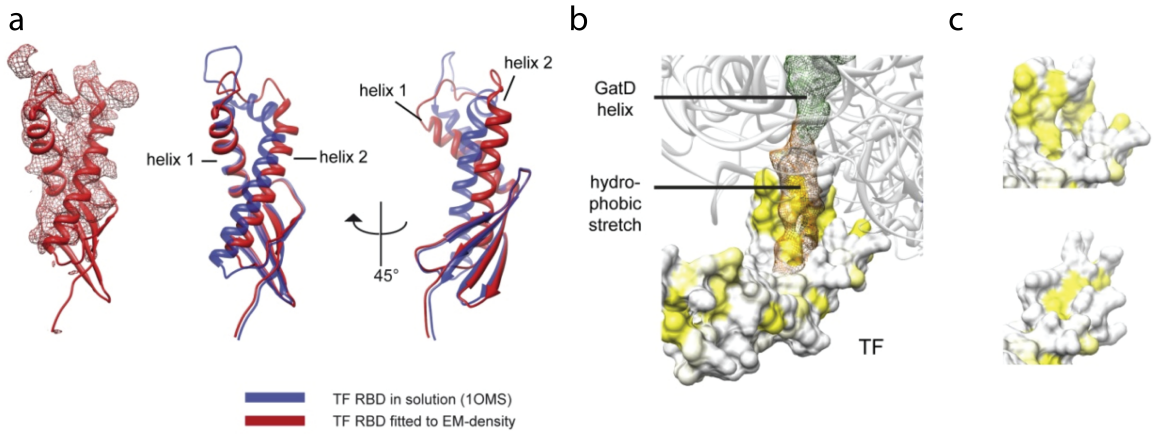

Figure 6.2: (a) A crystal structure of *E. coli* TF RBD in solution (1OMS [180]) was fitted into the EM density and regularized using the Coot algorithm [181]. The electron density is represented as a red mesh with the ribbon representation for the fitted model in solid red in the left panel. The middle panel shows a comparison of the crystal structure of TF RBD in solution (1OMS, blue) with the TF RBD model fitted to the EM density (red). The right panel shows the same comparison after rotation by 45° around the y-axis. (b) Surface hydrophobicity (yellow) of the fitted TF RBD with electron density for hydrophobic stretch (yellow) and alpha helix (green) of GatD NC. (c) Comparison of surface hydrophobicity representations of the ribosome bound TF RBD model fitted to the EM density (top) and the TF RBD crystal structure in solution (1OMS, bottom).

Conformational rearrangement of RBD upon ribosome binding. For a molecular interpretation of the density of the RBD we attempted to rigidly fit the crystal structure of *E. coli* TF RBD in solution (PDB: 1OMS [180]) into the density. Even though the two characteristic α-helices (1 and 2) of the TF RBD were clearly resolved in our density map, the crystal structure did not fit well by rigid body docking (Fig. 6.2a). To explain our density, helix 1 and helix 2 were manually adjusted by shifts into the electron density using the Coot software [181] (Fig. 6.2a). The rearranged helices are localized directly where the hydrophobic stretch of GatD first encounters the RBD of trigger factor (see Fig. 6.2b). The rearrangement of the two helices caused a presentation of a much larger hydrophobic surface (see Fig. 6.2c), compared to that of the TF RBD in solution (PDB: 1OMS), favoring interactions with the hydrophobic stretch of GatD. The possibility that such a rearrangement might lead to the formation of a hydrophobic "landing surface" for the nascent chain was already suggested [177] and is in agreement with chemical cross-link data [178].

Flexibility of TF decreases with longer nascent peptide. A recent study reported that *in vivo* TF engages ribosomes stable enough for detection only after translation of about 100 amino acids [179]. We therefore wondered whether the high degree of flexibility in our TF structure was due to the short length of our nascent GatD chain (85 amino acids). To address this question we prepared another sample of GatD-RNCs, where we extended the nascent chain to a total length of 145 amino acids (GatD-RNCs^{145}).

First, we performed microscale thermophoresis analysis and revealed that the affinity of TF for binding to GatD-RNC^{145} was fivefold higher than for binding to GatD-RNC^{85}. Second, we determined a second cryo-EM reconstruction of TF in complex with RNCs carrying the longer GatD^{145} nascent chain at 7.8 Å resolution. In contrast to the first structure, more of TF was visible. Low pass filtered maps (10.5 Å) displayed sufficient density to accommodate full length trigger factor indicating indeed a less flexible state of the head domain in this complex. Earlier crosslinking studies had already shown that nascent chains of at least 90 AA length can be crosslinked to the head domain of TF [178]. Therefore, we speculated that this second GatD nascent chain would be long enough to interact with the PPIase domain since it provides more possible interaction sites between TF and the nascent chain, and thereby might increase the binding affinity of TF to RNC and reduce the flexibility of TF. Our speculation is confirmed by MD simulations (see MD section).

Further sub-sorting of the EM images resulted in two sub-populations that differed in the conformation of the head domain. Whereas in conformation 1 one of the arms could make a contact to helix 24 of the 23S rRNA and the head domain to helix 98 (Fig. 6.3), in conformation 2 the head domain seems to contact helix 19 of the 23S rRNA (Fig. 6.3).

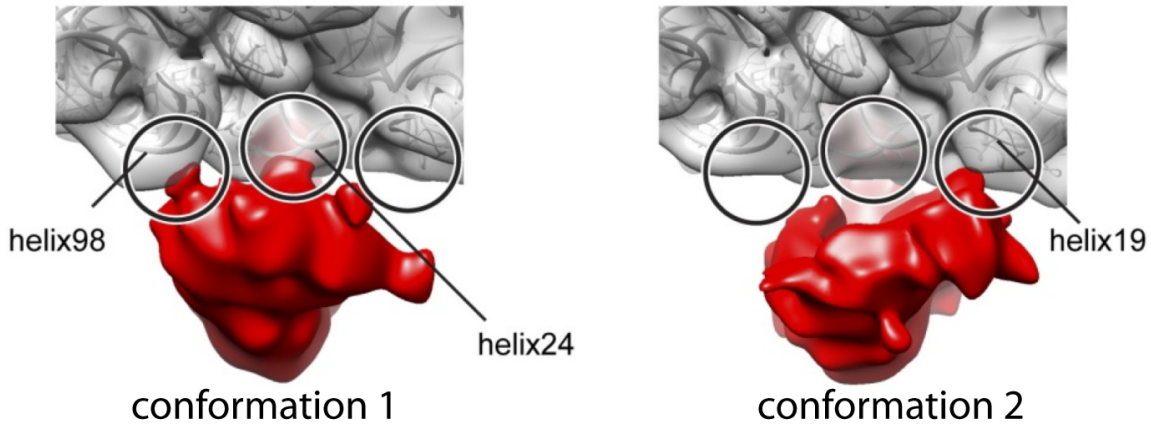

Figure 6.3: Cryo-EM reconstruction of the 70S-GatD-RNC-TF complex carrying the longer NC (filtered at 13.1 Å for clarity). A molecular model (2WWQ [13]) of the 50S subunit is fitted as a ribbon representation into the electron density map shown in grey. With the longer nascent chain, the PPIase domain of TF adopts two different conformations. Possible contacts of trigger factor to the ribosomal 50S subunit are indicated in both conformations.

Molecular dynamics simulations of the RNC-bound trigger factor

An atomic model of the ribosome-trigger factor-nascent chain complex (Fig. 6.1) was built by applying the molecular dynamics flexible fitting (MDFF) method [3, 4] to EM data of the GatD-RNC^{85} complex (see Methods). Based on this atomic model, two smaller sub-systems were constructed for simulations

(see Methods), one with a nascent GatD chain of 85 amino acids and one with a chain of 145 amino acids. Four independent simulations, each 1.2 μs long, for both sub-systems were performed using PACE, a hybrid united-atom/coarse-grained force field [182, 183, 184, 185, 186]. The simulations with the 85-AA and the 145-AA nascent chain are referred to as the short chain simulations and the long chain simulations, respectively.

Inter-domain motions of TF via pivot points. To characterize the motions of TF in the simulations, we performed principal component analysis (PCA), using the *ProDy* software [187], to reveal the dominant modes of TF motions. In the short chain simulations, the first two PCA modes represent inter-domain motions between the RBD and the other two domains (Fig. 6.4) via the first pivot point (residue 111 to 121). The third PCA mode captures a relative motion between the C-terminal and the head domain (Fig. 6.4) via the second pivot point (residue 149 to 154 and residue 241 to 248). For the long chain simulations, similar inter-domain motions via the two pivot points are observed in the first two PCA modes (Fig. 6.4). These characteristic inter-domain motions via pivot points are in agreement with reported MD simulations of TF monomer in solution [188, 189].

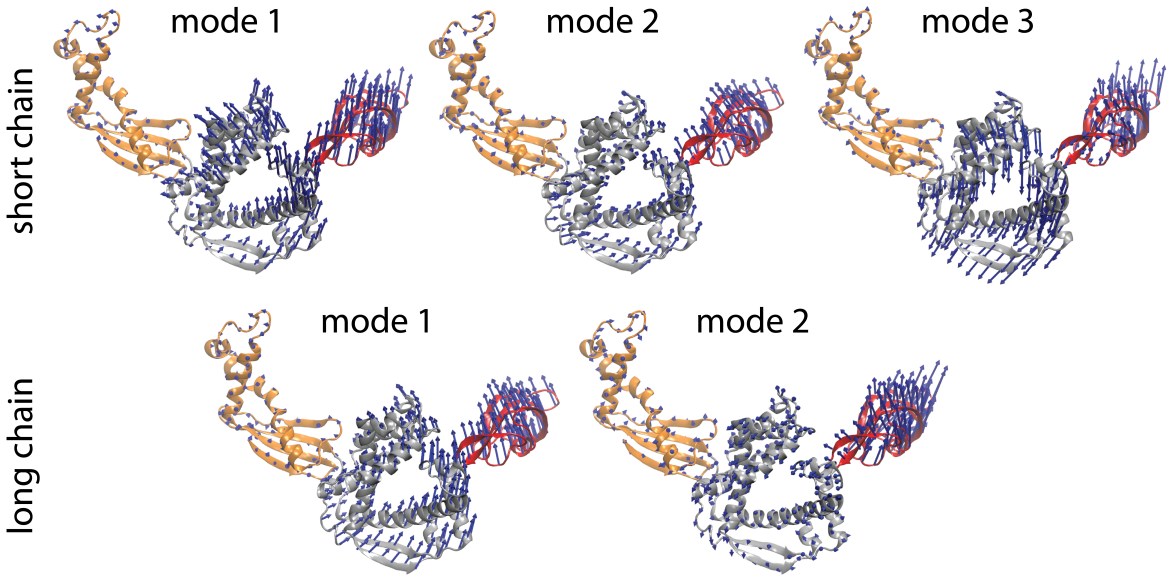

Figure 6.4: Principal component analysis of TF motions in the simulations. The top panel shows the first three PCA modes for the short chain simulations and the bottom panel shows the first two PCA modes for the long chain simulations. The RBD, the C-terminal domain and the head domain is colored orange, grey and red, respectively. The blue arrows represent the motions. Only the first 500 ns of the simulations were analyzed by PCA to exclude the bound and stable conformations. The simulations were aligned by the RBD before PCA was performed. These modes capture the characteristic inter-domain motions of the TF via pivot points.

Conformations of trigger factor bound with nascent chains of different length. In the short chain simulations, the head domain of the trigger factor binds to the ribosome surface (Fig. 6.5a) within a microsecond. The trigger factors span the ribosome surface with widespread, extended conformations (Fig. 6.5c). The wide range of binding conformations adopted by the ribosome-bound trigger factors in simulations agrees with EM experiments that the trigger factor bound to the 85-AA RNC is poorly resolved beyond the RBD due to its conformational variability.

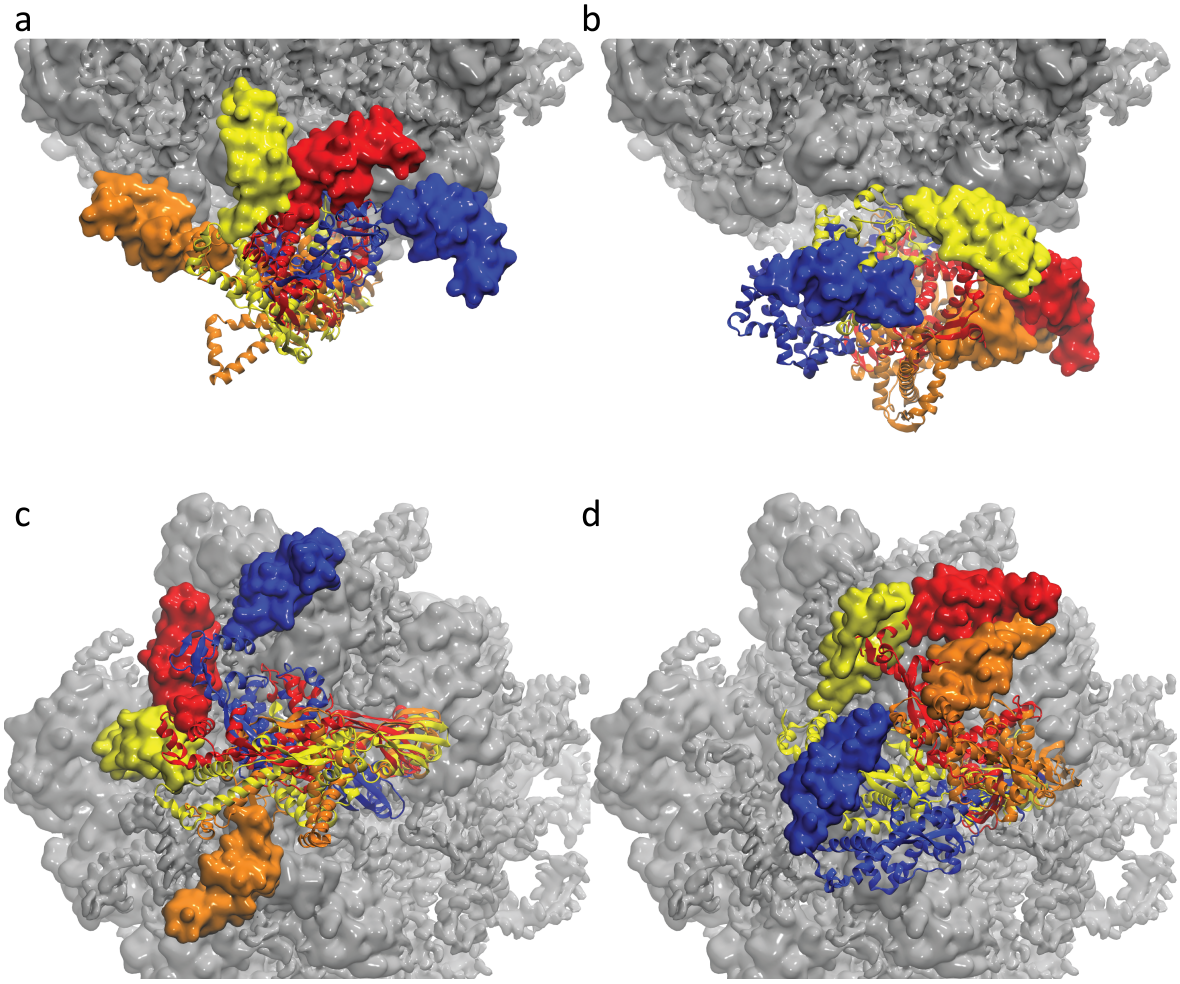

Figure 6.5: Average conformation of the trigger factor in the last 200 ns of the 85-AA chain simulations (a and c) and the 145-AA chain simulations (b and d). Figure a and b show the side view and figure c and d show the bottom view of the system. The trigger factor is colored blue, red, orange and yellow, respectively for each of the four simulations, with molecular surface of the head domain shown for clarity. The ribosome is colored grey. The trigger factor is further away from the ribosome (comparing a with b) and adopts a more compact conformation (comparing c with d) in the 145-AA chain simulations.

In contrast, the head domain of the trigger factor is further away from the ribosome (Fig. 6.5b) in the long chain simulations. Averaged over 800 ns (last 200 ns of the four simulations), the height of the center of mass of the head domain relative to the ribosome surface is 29.6 Å and 14.4 Å in the

long and short chain simulations, respectively, supporting the observed lifting of the trigger factor in the 145-AA chain simulations. Moreover, the trigger factor adopts more compact conformation in the long chain simulations (Fig. 6.5d). The radius of gyration of the trigger factor, averaged over 800 ns (last 200 ns of the four simulations), was determined to be: 30.6 Å in the long chain simulations *vs.* 33.6 Å in the short chain simulations, confirming the difference in the compactness of trigger factor structure in the two sets of simulations. In the EM experiments, the trigger factor is better resolved when bound with the 145-AA nascent peptide, which is consistent with the more compact structure of the trigger factor observed in the simulations with the 145-AA chain.

We determined the percentage of time in which the C-terminal and the head domain are in contact with the ribosome in the simulations. As shown in Table 6.1, the two TF domains (C-terminal and head) and the ribosome are frequently in contact in the short chain simulations. In three of the long chain simulations, the two domains contact the ribosome much less frequently. However in the remaining one long chain simulation, the C-terminal and the head domain are in contact with the ribosome in 28.5% and 35.8% of the simulation time, respectively. Moreover, the final TF conformation in this particular simulation overlaps indeed partially with the EM densities of the proposed conformation 2 of the GatD-RNC^{145} complex (Fig. 6.6). Even though the head domain in this simulated conformation does not contact helix 19 directly, it is in close proximity to the densities of the proposed contact region (Fig. 6.6). Therefore it is possible that the head domain will reach the proposed contact region and the proposed conformation 2 will be reproduced in the simulations if more sampling is given.

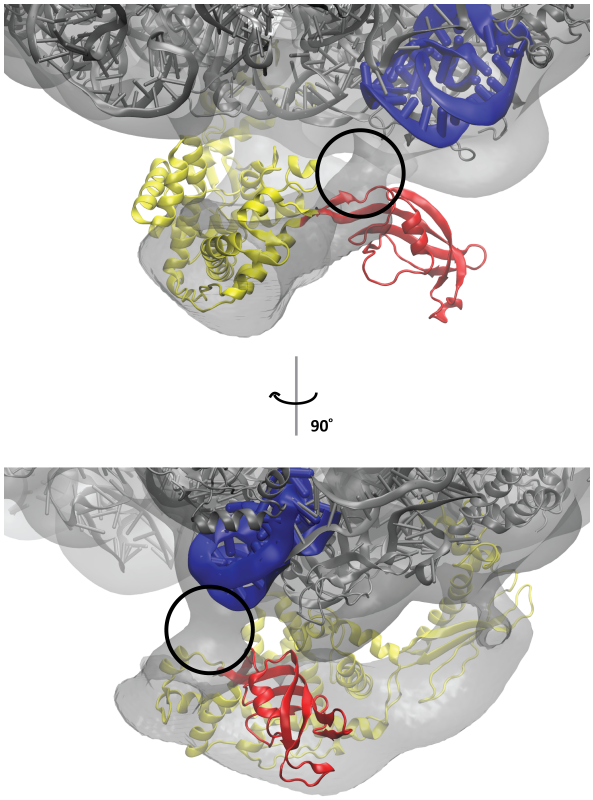

Figure 6.6: Final stable conformation of TF in one of the long chain simulations is shown (in two views) inside the EM map of conformation 2 of the GatD-RNC^{145} complex (filtered at 20 Å). The trigger factor is colored yellow for the RBD and the C-terminal domain and red for the head domain. The ribosome is shown in grey with helix 19 in blue. The proposed contact region between the head domain and helix 19 is circled. This TF conformation overlaps partially with the EM densities and the head domain is in close proximity to the proposed contact region.

Table 6.1: Percentage of time in which the C-terminal domain and the head domain are in contact with the ribosome in the simulations. A TF domain (C-terminal or head) is considered in contact with the ribosome when the heavy atoms of the TF domain and the ribosome are within 4 Å in distance.

	85-AA chain simulations				145-AA chain simulations			
C-terminal domain - ribosome	0.2%	45.0%	6.3%	19.2%	0.0%	0.2%	0.0%	28.5%
head domain - ribosome	63.8%	31.8%	22.7%	47.2%	20.5%	8.2%	4.0%	35.8%

Interactions with the nascent peptide. In the short chain simulations, molecular contacts between the 85-AA chain and the head domain are few (Fig. 6.7a and Table 6.2), therefore the head domain is not heavily restricted in motion by the nascent peptide and exhibits large conformational variability. In contrast, in the long chain simulations, the N-terminus of the chain forms a "collapsed coil" structure which establishes extensive contacts with the head domain (Fig. 6.7b and Table 6.2). The higher binding affinity of TF to RNC with the 145-AA nascent chain observed in microscale thermophoresis experiments can be attributed to the more extensive interactions between the TF head domain and the 145-AA chain. Moreover, these interactions pull the head domain further away from the ribosome surface and keep the head domain close to the C-terminal domain, resulting in a more compact structure of the trigger factor.

Table 6.2: Average number of molecular contacts between the nascent chain and the head domain of the trigger factor in the last 200 ns of the simulations. Two residues are considered in contact when their C_α atoms are within 6 Å in distance. The left and right column shows the number for the four short chain simulations and the four long chain simulations, respectively. More contacts are observed between the head domain and the 145-AA chain in the simulations.

85 AA	145 AA
13.17	29.54
13.50	25.35
5.10	33.37
0	35.86

The types of interactions between the nascent peptide and the trigger factor in the simulations were monitored. The RBD of the trigger factor interacts with the nascent peptide mainly through hydrophobic interactions (Fig. 6.7c and Appendix B Table B.1 and B.2), agreeing with the known hydrophobic binding motif in the RBD [190]. Hydrophobic interactions still dominate between the C-terminal domain and the nascent peptide, but more hydrophilic interactions arise (Fig. 6.7c and Appendix B Table B.3 and B.4), showing that the trigger factor can interact with, and hence offer protection to, nascent peptides of various hydrophobicity [178]. For the head domain, only a few residues are found interacting with the nascent peptide in the 85-AA chain simulations (Appendix B Table B.5), while much more residues interact with the 145-AA nascent chain through both hydrophobic and hydrophilic interactions (Fig. 6.7c and Appendix B Table B.6).

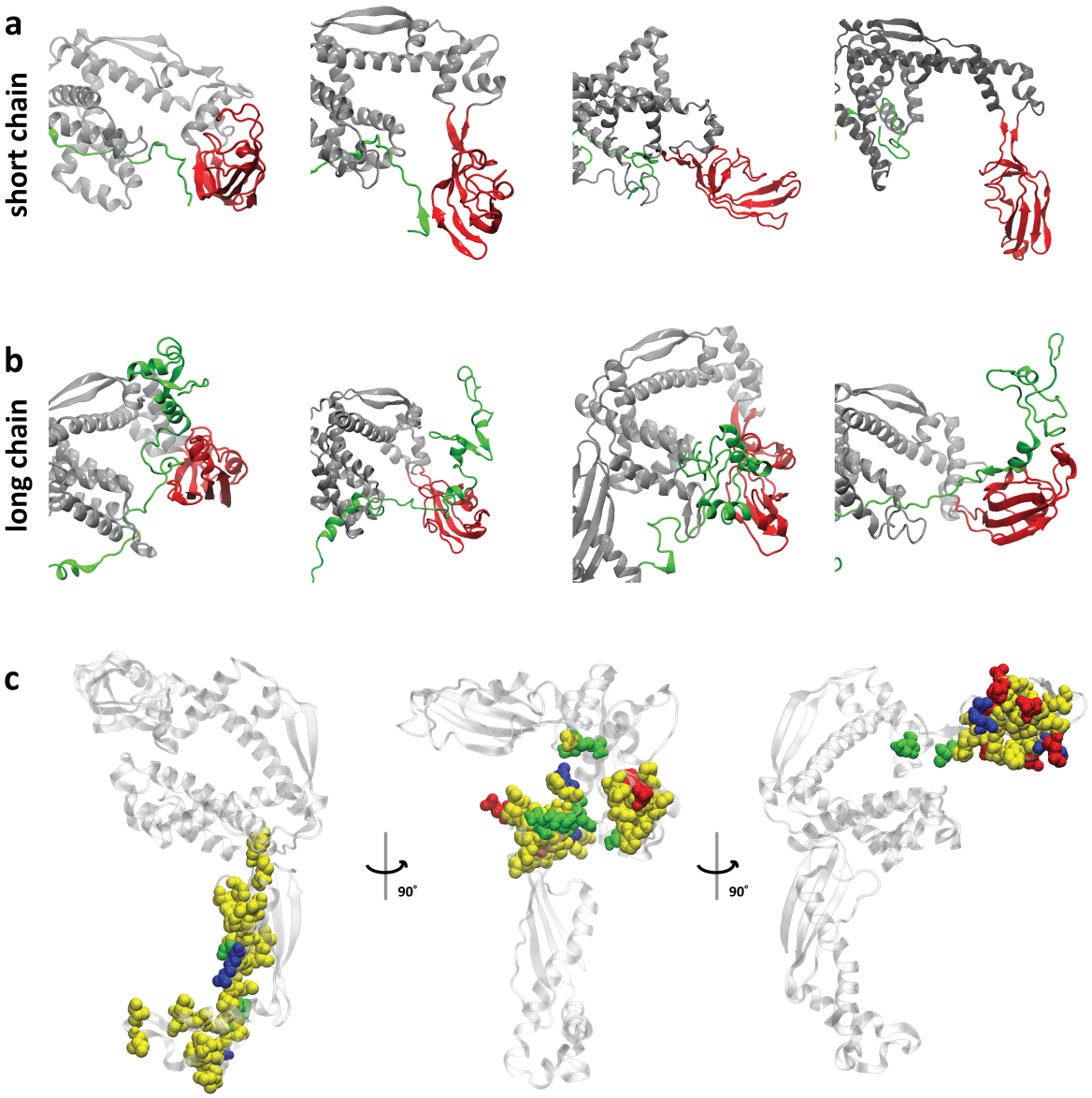

Figure 6.7: (a) Average conformation of the N-terminus of the 85-AA nascent peptide in the last 200 ns of the simulations. The nascent peptide is colored green and the trigger factor is colored red for the head domain and grey for the C-terminal domain. The N-terminus of the nascent chain contacts slightly with or even detaches from the head domain of the trigger factor. (b) Average conformation of the N-terminus of the 145-AA nascent peptide in the last 200 ns of the simulations. The same coloring scheme as in (a) is employed. The N-terminus of the 145-AA nascent chain forms a "collapsed coil" structure which contacts extensively with the head domain of the trigger factor. (c) Residues of trigger factor in the ribosome binding domain (left), the C-terminal domain (middle) and the head domain (right) that interact with the nascent peptide in the simulations are shown. These residues are shown with a VdW representation according to the following coloring scheme: residues that participate in hydrophobic interactions with the chain are colored yellow while residues that interact through hydrophilic interactions are colored blue, red and green for positively charged, negatively charged and neutral amino acids, respectively.

6.3 Discussion

We employed microseconds-long MD simulations to study the dynamics of the ribosome-bound trigger factor. Our results reveal the different conformational variability of the trigger factor when bound

with a short and a long nascent peptide, explaining why TF is better resolved when bound with a longer nascent chain in the EM experiments. We showcased that MD simulations can complement cryo-EM analyses to characterize the dynamics of molecular complexes.

Moreover, this work offers insight into the chaperone function of the trigger factor. Trigger factor functions by forming a molecular cradle over the exit of the ribosomal exit tunnel to offer protection to the nascent peptide. We revealed that, upon binding to the ribosome, the RBD of the TF undergoes conformational rearrangement to expose a larger hydrophobic surface for interacting with the exiting nascent peptide. Moreover, as the nascent chain matures by elongation, the TF head domain is capable of forming more extensive interactions with the longer nascent peptide as shown in the simulations, explaining the higher binding affinity of TF to RNC with a longer nascent chain. These two mechanisms enhance the TF-nascent chain interactions, and thereby, assist the recruitment of TF to the RNC, which is essential for the chaperone function of the trigger factor.

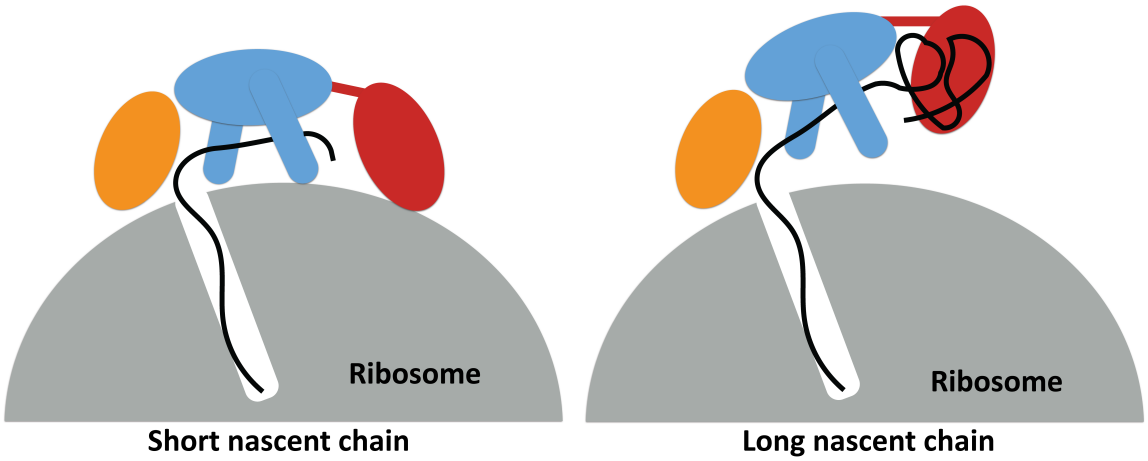

Figure 6.8: Schematic figure illustrating how the length of the nascent peptide modulate the conformation of the trigger factor. The ribosome is colored grey; the RBD, the C-terminal domain and the head domain of the trigger factor is colored orange, blue and red, respectively. (A) With a short nascent chain, the head domain and the nascent chain do not have much interactions and the head binds to the ribosome. (B) With a long nascent chain, the N-terminus of the chain forms a "collapsed coil" structure which binds with the head domain, lifting the head domain and pulling it close to the C-terminal domain.

TF binds to the ribosome-nascent chain complex and as elongation of the nascent peptide proceeds, it leaves the ribosome while staying bound to the nascent peptide, allowing a next trigger factor to be recruited to protect the elongating chain [191]. Moreover, the length of the nascent peptide has been shown to modulate the dissociation of the trigger factor from the ribosome [192], but the molecular mechanism is not clearly known. The simulations offer a molecular explanation: as the nascent peptide grows longer, the trigger factor is lifted further away from the ribosome by the molecular interactions

between the head domain and the longer nascent peptide (see Fig. 6.8). The lifting could facilitate the unbinding of the trigger factor from the ribosome. These interactions also serve to keep the trigger factor bound to the nascent chain after it leaves the ribosome.

6.4 Methods

Cryo-EM and microscale thermophoresis experiments

Experimental details of the cryo-EM and the microscale thermophoresis experiments are described in [9].

MDFF modeling

The atomic model of the ribosome-trigger factor-nascent chain complex was built in steps. First, an atomic structure of the ribosome [13] (PDB: 2WWQ-2WWL) was flexibly fit into the ribosomal densities of the 7.7 Å EM map of the GatD-RNC^{85} complex by the molecular dynamics flexible fitting (MDFF) method [3, 4]. Second, the structure of the nascent chain inside the exit tunnel was modeled with HHpred [193], and then refined by MDFF simulation with a positionally constrained ribosome. Third, the trigger factor was modeled starting from a crystal structure of the trigger factor (PDB: 1W26 [167]), with helix 1 and 2 of the RBD rearranged by Coot as described in Results. The RBD, the C-terminal domain and the head domain were then fitted into three EM maps of the GatD-RNC^{85} complex, resolved at 7.7 Å, 16.5 Å and 35 Å resolution, respectively, in one MDFF simulation (Fig. 6.1). Each domain was coupled only to the respective EM map in the MDFF simulation. Since the C-terminal and the head domain were fitted to maps of lower resolution, the secondary structure restraints [3] applied to the C-terminal domain were twice stronger than default, and domain restraints, which keep the backbone structure rigid during fitting, were applied to the head domain. This way of performing MDFF, namely coupling various domains to maps of different resolutions with different level of structural restraints, permits EM data at lower resolution to be employed for flexible fitting. To improve the quality of the fitted models, a better solvent description, compared to that in in-vacuo fitting [3], was furnished by the generalized Born implicit solvent model [6] in NAMD [11] during all the MDFF simulations. All MDFF simulations were performed for 5 ns, using the same set of parameters as employed in [6].

Simulated systems and molecular dynamics protocols

To carry out the hybrid simulations, the nascent chain needs to be extended beyond the exit tunnel. Without well-defined EM densities for the nascent chain outside the ribosome, an initial model of the 85-AA nascent peptide was built by placing amino acids inside the cradle of the trigger factor according to distance information from cross-linking experiments [178]. The chain was then subjected to 50 ns of all-atom equilibrium simulations in implicit solvent [33], with the TF and the ribosome constrained in positions and with harmonic distance restraints applied between the nascent chain and the trigger factor; the distance restraints were set up based on the cross-linking experiments [178]. To extend the chain to 145 amino acids, the model of the 85-AA chain was mutated and 60 more amino acids were added in a straight conformation to the N-terminus, followed by 50 ns of all-atom equilibration simulations in implicit solvent [33]. Position and distance restraints similar to those applied in the equilibration of the 85-AA chain were employed.

The two atomic models were cut smaller to reduce the simulations size by retaining only molecular structures that are within 80 Å of the trigger factor. Ribosome-bound Mg^{2+} ions were added by the cionize [194] plugin of VMD [116]. The sub-systems were then coarse-grained and described at hybrid resolutions using the PACE force field [183, 195, 185]. This force field, which was developed as a complement to the MARTINI force field [196], describes the solvent environment at a coarse-grained level while representing proteins at a united-atom level, allowing efficient long-time simulations of folding of several proteins into their native structures [184, 185, 186]. In the present study, the ribosomal proteins, the trigger factor and the nascent peptide were modeled at a united-atom level. The ribosomal RNAs were described with a coarse-grained RNA model [197] developed recently on the basis of the MARTINI force field. The rest of the systems, including water and ions, were modeled by means of the MARTINI force field. Each system was solvated in a 220 x 220 x 220 $Å^3$ water box and buffered at 0.1 M Na^+Cl^- concentration. All simulations were performed in NPT ensemble using a version of NAMD 2.9 [11] which supports the PACE force field. The following parameters were employed in the simulations: temperature was maintained at 310 K using Langevin dynamics with a damping constant of 0.2 ps^{-1}; pressure was maintained at 1 atm using Nosè-Hoover Langevin piston pressure control with a oscillation time of 100 fs and a decay time of 50 fs. the RESPA multiple-time-stepping algorithm [81, 82] was adopted with a 4-fs integration time step, short-range forces calculated every time step, and long-range electrostatics evaluated every time step. During the simulations, the ribosomal RNA beads and the ribosome-bound Mg^{2+} ions were fixed in positions,

which was justified as such constraints have been shown to have only a small effect on simulated nascent chain dynamics [198, 199, 200].

Chapter 7

Conclusions and outlook

With the continuous advance in instrumentation hardware as well as 3D image reconstruction methodology, cryo-EM can now routinely resolve EM structures at sub-nanometer resolution. Hybrid approaches greatly assist the interpretation of EM data by furnishing atomic models of the imaged systems. Structure determination of macromolecular complexes is largely benefited by, and at the same time pushes the development of, cryo-EM imaging and hybrid methods, as exemplified by the ribosome. In this thesis works we focused on the methodological advancement of the MDFF hybrid method and the applications to the ribosome. Structural symmetry can now be incorporated into the MDFF protocol through symmetry-restrained MDFF [5]. Use of the generalized Born implicit solvent model overcomes shortcomings of in vacuo MDFF simulations [6]. Participation in the Cryo-EM Modeling Challenge 2010 showcased the performance of MDFF against other hybrid approaches [7]. For applications to the ribosome, MDFF model of the Tet(O)-bound ribosomal complex revealed structural elements of Tet(O) that are critical to Tet(O)-mediated bacterial resistance to the tetracycline antibiotic [8]. Finally a combination of MDFF and subsequent microseconds-long MD simulations complemented a cryo-EM study about the dynamics of the ribosome-bound trigger factor [9], which are shown to be modulated by the nascent polypeptide.

The works in this thesis indeed demonstrated several intrinsic advantages of the MDFF method. First, the extensive set of features available in NAMD [11], the underlying simulation package of MDFF, allows inclusion of many different forms of structural restraints into the fitting simulations. Therefore users' experience and knowledge about the structure can be easily and flexibly incorporated into fitting protocol, as showcased in symmetry-restrained MDFF [5] and our participation in the Cryo-EM Modeling Challenge 2010 [7]. Second, since MDFF is based on MD simulations, a realistic environment can be furnished by an explicit or implicit solvent model [6], or even by a lipid membrane [7], to improve the quality of fitted models. Third, as shown in the study of Tet(O)-mediated bacterial resistance to antibiotics, homology modeling can be employed jointly with MDFF to perform structural analyses on systems without crystallographic structures. Lastly, the readiness of MDFF

models for further MD simulations permits the dynamics of molecular complexes be characterized by MD in conjunction with cryo-EM analyses, as exemplified by our study of the trigger factor [9].

Looking forward, two aspects of cryo-EM modeling by MDFF should be further enhanced. First, a closer blend between MDFF and homology modeling or structure prediction softwares is desired. As cryo-EM imaging is continuously applied to molecular complexes of larger size, such as the recently resolved HIV capsid structure [201], and of higher level organisms, such as the EM structure of the human ribosome [202], it is highly likely that crystallographic structures of certain components of the system of interest are missing. Homology modeling or structure prediction software will be needed to provide initial structures for these components in order to perform MDFF. Therefore a good interfacing with these softwares, or even a united algorithm combining MDFF with these softwares, will greatly assist the use of MDFF. Second, a **quantitative** matching between the dynamic information extracted from EM data and from MD simulations is of great interest. In particular, it has been proposed that motion-affected EM densities can be interpreted in terms of MD trajectories. In the study of the trigger factor, microseconds-long simulations provide a **qualitative** description of the dynamics of the trigger factor to complement cryo-EM analyses. However with the present MD technique and computing power, the conformational sampling achieved by MD is still insufficient to provide a **quantitative** interpretation of motion-affected EM densities. In spite of that, the combination of MDFF and subsequent MD simulations is still a promising approach to interpret dynamic information in EM data, as MD is a natural choice to reproduce the ensemble of conformations captured in EM. Further advance of MD methodology, such as coarse-grained MD and enhanced sampling technique, will be needed though to resolve the sampling issue.

Appendix A

Theoretical description of the generalized Born implicit solvent model

The GBIS model [86] represents polar solvent as a dielectric continuum and, accordingly, screens electrostatic interactions between solute atoms. GBIS treats solute atoms as spheres of low protein dielectric ($\epsilon_{\rm p} = 1$), whose radius is the Bondi [203] van der Waals radius, in a continuum of high solvent dielectric ($\epsilon_{\rm s} = 80$).

The total electrostatic energy for atoms in a dielectric solvent is modeled as the sum of Coulomb and generalized Born (GB) energies [86],

$$E_{\rm T}^{\rm Elec} \;=\; E_{\rm T}^{\rm Coul} + E_{\rm T}^{\rm GB} \;. \tag{A.1}$$

The total Coulomb energy for the system of atoms is the sum over pairwise Coulomb energies,

$$E_{\rm T}^{\rm Coul} \;=\; \sum_{i} \sum_{j>i} E_{ij}^{\rm Coul} \;, \tag{A.2}$$

where the double summation represents all unique pairs of atoms within the interaction cutoff; the interaction cutoff for GBIS simulations is generally in the range 16-20 Å, i.e., longer than for explicit solvent simulations, where it is typically 8-12 Å. The reason for the wider cutoff is that particle-mesh Ewald summations, used to describe long-range Coulomb forces, cannot be employed for treatment of long-range GBIS electrostatics.

The pairwise Coulomb energy, $E_{ij}^{\rm Coul}$ in eq. A.2, is

$$E_{ij}^{\rm Coul} \;=\; (k_{\rm e}/\epsilon_{\rm p})\, q_i q_j / r_{ij} \;, \tag{A.3}$$

where $k_{\rm e} = 332$ (kcal/mol)Å/e^2 is the Coulomb constant, q_i is the charge on atom i, and r_{ij} is the distance between atoms i and j. The total GB energy for the system of atoms is the sum over pairwise

GB energies and self-energies given by the expression

$$E_{\mathrm{T}}^{\mathrm{GB}} \;=\; \underbrace{\sum_{i}\sum_{j>i} E_{ij}^{\mathrm{GB}}}_{\text{pair}} + \underbrace{\sum_{i} E_{ii}^{\mathrm{GB}}}_{\text{self}} \,, \tag{A.4}$$

where the pair-energies and self-energies are defined as [86]

$$E_{ij}^{\mathrm{GB}} \;=\; -\,(k_{\mathrm{e}} D_{ij})\, q_i q_j / f_{ij}^{\mathrm{GB}} \,. \tag{A.5}$$

Here, D_{ij} is the pairwise dielectric term [204], which contains the contribution of an implicit ion concentration to the dielectric screening, and is expressed as

$$D_{ij} \;=\; (1/\epsilon_{\mathrm{p}}) - \exp\left(-\kappa f_{ij}^{\mathrm{GB}}\right)/\epsilon_{\mathrm{s}} \,, \tag{A.6}$$

where κ^{-1} is the Debye screening length which represents the length scale over which mobile solvent ions screen electrostatics. For an ion concentration of 0.2 M, room temperature water has a Debye screening length of $\kappa^{-1} = \sim$7 Å. f_{ij}^{GB} is [86]

$$f_{ij}^{\mathrm{GB}} \;=\; \sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp\left(-r_{ij}^2/4\alpha_i\alpha_j\right)} \,. \tag{A.7}$$

The form of the pairwise GB energy in eq. A.5 is similar to the form of the pairwise Coulomb energy in eq. A.3, but is of opposite sign and replaces the $1/r_{ij}$ distance dependence by $1/f_{ij}^{\mathrm{GB}}$. The GB energy bears a negative sign because the electrostatic screening counteracts the Coulomb interaction. The use of f_{ij}^{GB}, instead of r_{ij}, in eq. A.5 heavily screens the electrostatic interaction between atoms which are either far apart or highly exposed to solvent. The more exposed an atom is to high solvent dielectric, the more it is screened electrostatically, represented by a smaller Born radius, α_i.

Accurately calculating the Born radius is central to a GBIS model as the use of perfect Born radii allows the GBIS model to reproduce, with high accuracy, the electrostatics and solvation energies described by the Poisson-Boltzmann equation [205], and does it much faster than a Poisson-Boltzmann or explicit solvent treatment [206]. Different GBIS models vary in how the Born radius is calculated; models seek to suggest computationally less expensive algorithms without undue sacrifice in accuracy. Many GBIS models [207] calculate the Born radius by assuming atoms are spheres whose radius is the Bondi [203] van der Waals radius and determine an atom's exposure to solute through the sum of

overlapping surface areas with neighboring spheres [208]. The more recent GBIS model of Onufriev, Bashford and Case ($\mathrm{GB}^{\mathrm{OBC}}$), applied successfully to MD of macromolecules [209, 210] and adopted in NAMD, calculates the Born radius as

$$\alpha_i \;=\; \left[(1/\rho_{i0}) - (1/\rho_i)\tanh\left(\delta\psi_i - \beta\psi_i^2 + \gamma\psi_i^3\right)\right]^{-1} \;, \tag{A.8}$$

where ψ_i, the sum of surface area overlap with neighboring spheres, is calculated through

$$\psi_i \;=\; \rho_{i0}\sum_j H(r_{ij},\rho_i,\rho_j) \;. \tag{A.9}$$

As explained in prior studies [208, 207, 210], $H(r_{ij},\rho_i,\rho_j)$ is the surface area overlap of two spheres based on their relative separation, r_{ij}, and radii, ρ and ρ_0; the parameters δ, β and γ in eq. A.8 have been calculated to maximize agreement between Born radii described by eq. A.8 and those derived from Poisson-Boltzmann electrostatics [210].

The total electrostatic force acting on an atom is the sum of Coulomb and GB forces; the net Coulomb force on an atom is given by

$$\vec{F}_i^{\mathrm{Coul}} \;=\; -\sum_j \left[dE_{\mathrm{T}}^{\mathrm{Coul}}/dr_{ij}\right]\hat{r}_{ij} \;. \tag{A.10}$$

whose derivative $(dE_{\mathrm{T}}^{\mathrm{Coul}}/dr_{ij})$ is inexpensive to calculate. The required derivatives $(dE_{\mathrm{T}}^{\mathrm{GB}}/dr_{ij})$ for the GB force, however, are much more expensive to calculate because E_{ij}^{GB} depends on interatomic distances, r_{ij}, both directly (c.f. eqs. A.5 and A.7) and indirectly through the Born radius (c.f. eqs. A.5, A.7, A.8 and A.9). The net GB force on an atom is given by

$$\begin{aligned}
\vec{F}_i^{\mathrm{GB}} &= -\sum_j \left[dE_{\mathrm{T}}^{\mathrm{GB}}/dr_{ij}\right]\hat{r}_{ij} \\
&= -\sum_j \left[\sum_k\sum_{l>k}(\partial E_{\mathrm{T}}^{\mathrm{GB}}/\partial r_{kl})(dr_{kl}/dr_{ij}) + \sum_k(\partial E_{\mathrm{T}}^{\mathrm{GB}}/\partial\alpha_k)(d\alpha_k/dr_{ij})\right]\hat{r}_{ij} \\
&= -\sum_j \left[\partial E_{\mathrm{T}}^{\mathrm{GB}}/\partial r_{ij} + (\partial E_{\mathrm{T}}^{\mathrm{GB}}/\partial\alpha_i)(d\alpha_i/dr_{ij}) + (\partial E_{\mathrm{T}}^{\mathrm{GB}}/\partial\alpha_j)(d\alpha_j/dr_{ij})\right]\hat{r}_{ij} \;,
\end{aligned} \tag{A.11}$$

with $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$. The required partial derivative of $E_{\mathrm{T}}^{\mathrm{GB}}$ with respect to a Born radius, α_k, is

$$\partial E_{\mathrm{T}}^{\mathrm{GB}}/\partial\alpha_k = \sum_i\sum_{j>i}\left[\partial E_{ik}^{\mathrm{GB}}/\partial\alpha_k + \partial E_{kj}^{\mathrm{GB}}/\partial\alpha_k\right] + \sum_i \partial E_{ii}^{\mathrm{GB}}/\partial\alpha_k \;. \tag{A.12}$$

The summations in eqs. A.9, A.11 and A.12, require three successive iterations over all pairs of atoms for each GBIS force calculation, whereas calculating Coulomb forces for an explicit solvent simulation requires only one such iteration over atom-pairs. Also, because of the computational complexity of the above GBIS equations, the total cost of calculating the pairwise GBIS force between pairs of atoms is $\sim 7\times$ higher than the cost for the pairwise Coulomb force. For large systems and long cutoffs, the computational expense of implicit solvent simulations can exceed that of explicit solvent simulations; however, in this case, an effective speed-up over explicit solvent still arises due to faster conformational exploration as illustrated in Chapter 3. The trade-off between the speed-up of implicit solvent models and the higher accuracy of explicit solvent models is still under investigation. [206]

Appendix B

Interactions between trigger factor and nascent polypeptide

Residues of trigger factor that interact with the nascent peptide in the short and the long chain simulations are listed in the six tables below. A TF residue is considered interacting with the nascent chain if the hydrophobic/hydrophilic heavy atoms of the side chain are within 4 Å from that of the nascent peptide in more than 10% of time in the last 200 ns of four simulations, i.e., more than 80 ns.

Table B.1: Residues in the RBD of the trigger factor that interact with the 85-AA nascent peptide in the short chain simulations and the types of the interactions (hydrophobic, **hydrophilic** or both).

Residues	Interaction types
Leu32	hydrophobic
Val35	hydrophobic
Ala36	hydrophobic
Val39	hydrophobic
Val49	hydrophobic
Pro50	hydrophobic
Met51	hydrophobic
Ile53	hydrophobic
Val54	hydrophobic
Arg57	**hydrophilic**
Tyr58	both
Val62	hydrophobic
Gln64	**hydrophilic**
Val66	hydrophobic
Leu67	hydrophobic
Met71	hydrophobic
Ser72	**hydrophilic**
Arg73	**hydrophilic**
Phe75	hydrophobic
Ile76	hydrophobic
Ile79	hydrophobic
Ile80	hydrophobic
Ile84	hydrophobic
Pro86	hydrophobic
Ala89	hydrophobic
Pro90	hydrophobic
Tyr92	hydrophobic
Val93	hydrophobic
Pro94	hydrophobic
Tyr112	both

Table B.2: Residues in the RBD of the trigger factor that interact with the 145-AA nascent peptide in the long chain simulations and the types of the interactions (hydrophobic, **hydrophilic** or both).

Residues	Interaction types
Leu32	hydrophobic
Val35	hydrophobic
Ala36	hydrophobic
Ala39	hydrophobic
Ile41	hydrophobic
Phe44	hydrophobic
Val49	hydrophobic
Pro50	hydrophobic
Met51	hydrophobic
Ile53	hydrophobic
Val54	hydrophobic
Arg57	**hydrophilic**
Tyr58	hydrophobic
Ala60	hydrophobic
Ser61	**hydrophilic**
Val62	hydrophobic
Val66	hydrophobic
Met71	hydrophobic
Phe75	hydrophobic
Ile76	hydrophobic
Ile80	hydrophobic
Pro90	hydrophobic
Tyr92	hydrophobic
Val93	hydrophobic
Pro94	hydrophobic

Table B.3: Residues in the C-terminal domain of the trigger factor that interact with the 85-AA nascent peptide in the short chain simulations and the types of the interactions (hydrophobic, **hydrophilic** or both).

Residues	Interaction types	Residues	Interaction types
Pro113	hydrophobic	Glu335	**hydrophilic**
Gln147	**hydrophilic**	Leu336	hydrophobic
Gln148	**hydrophilic**	Phe337	hydrophobic
Ala149	hydrophobic	Ala341	hydrophobic
Pro303	hydrophobic	Arg344	**hydrophilic**
Ala305	hydrophobic	Val345	hydrophobic
Leu306	hydrophobic	Leu349	hydrophobic
Ile307	hydrophobic	Leu370	hydrophobic
Glu310	**hydrophilic**	Ile371	hydrophobic
Ile311	hydrophobic	Glu373	**hydrophilic**
Val313	hydrophobic	Met374	hydrophobic
Leu314	hydrophobic	Ala377	hydrophobic
Gln317	**hydrophilic**	Tyr378	both
Ala318	hydrophobic	Pro381	hydrophobic
Ala319	hydrophobic	Glu383	**hydrophilic**
Gln320	**hydrophilic**	Val384	hydrophobic
Arg321	**hydrophilic**	Ile385	hydrophobic
Phe322	hydrophobic	Phe387	hydrophobic
Asn325	**hydrophilic**	Tyr388	hydrophobic
Gln328	**hydrophilic**	Leu394	hydrophobic
Leu330	hydrophobic	Asn397	**hydrophilic**
Leu332	hydrophobic	Met398	hydrophobic
Pro333	hydrophobic	Val401	hydrophobic

Table B.4: Residues in the C-terminal domain of the trigger factor that interact with the 145-AA nascent peptide in the long chain simulations and the types of the interactions (hydrophobic, **hydrophilic** or both).

Residues	Interaction types
Gln147	**hydrophilic**
Ser309	**hydrophilic**
Val313	hydrophobic
Arg316	**hydrophilic**
Ala319	hydrophobic
Gln320	**hydrophilic**
Arg321	**hydrophilic**
Phe322	hydrophobic
Glu326	**hydrophilic**
Ala329	hydrophobic
Leu330	hydrophobic
Leu332	hydrophobic
Pro333	hydrophobic
Arg334	**hydrophilic**
Glu335	**hydrophilic**
Leu336	hydrophobic
Phe337	hydrophobic
Arg344	**hydrophilic**
Ile371	hydrophobic
Met374	hydrophobic
Tyr378	both
Glu379	**hydrophilic**
Asp380	**hydrophilic**
Pro381	hydrophobic
Val384	hydrophobic
Ile385	hydrophobic
Phe387	hydrophobic
Tyr388	hydrophobic
Leu394	hydrophobic
Asn397	hydrophobic

Table B.5: Residues in the head domain of the trigger factor that interact with the 85-AA nascent peptide in the short chain simulations and the types of the interactions (hydrophobic, **hydrophilic** or both).

Residues	Interaction types
Thr150	**hydrophilic**
Trp151	both
Asp162	**hydrophilic**
Asp184	**hydrophilic**

Table B.6: Residues in the head domain of the trigger factor that interact with the 145-AA nascent peptide in the long chain simulations and the types of the interactions (hydrophobic, **hydrophilic** or both).

Residues	Interaction types
Trp151	hydrophobic
Ile166	hydrophobic
Phe168	hydrophobic
Val172	hydrophobic
Phe177	hydrophobic
Glu178	**hydrophilic**
Lys181	**hydrophilic**
Ala182	hydrophobic
Phe185	hydrophobic
Val186	hydrophobic
Leu187	hydrophobic
Ala188	hydrophobic
Met189	hydrophobic
Arg193	**hydrophilic**
Met194	hydrophobic
Ile195	hydrophobic
Pro196	hydrophobic
Phe198	hydrophobic
Glu199	**hydrophilic**
Asp200	**hydrophilic**
Phe217	hydrophobic
Pro218	hydrophobic
Glu220	**hydrophilic**
Tyr221	both
Ala223	hydrophobic
Glu224	**hydrophilic**
Leu226	hydrophobic
Phe233	hydrophobic
Ile235	hydrophobic
Glu241	**hydrophilic**

References

[1] J. Poehlsgaard and S. Douthwaite. The bacterial ribosome as a target for antibiotics. *Nat. Rev. Microbiol.*, 3:870–881, 2005.

[2] K. Mitra and J. Frank. Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. *Annu. Rev. Biophys. Biomol. Struct.*, 35:299–317, 2006.

[3] Leonardo G. Trabuco, Elizabeth Villa, Kakoli Mitra, Joachim Frank, and Klaus Schulten. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*, 16:673–683, 2008.

[4] Leonardo G. Trabuco, Elizabeth Villa, Eduard Schreiner, Christopher B. Harrison, and Klaus Schulten. Molecular Dynamics Flexible Fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods*, 49:174–180, 2009.

[5] Kwok-Yan Chan, James Gumbart, Ryan McGreevy, Jean M. Watermeyer, B. Trevor Sewell, and Klaus Schulten. Symmetry-restrained flexible fitting for symmetric EM maps. *Structure*, 19:1211–1218, 2011.

[6] David E. Tanner, Kwok-Yan Chan, James Phillips, and Klaus Schulten. Parallel generalized Born implicit solvent calculations with NAMD. *J. Chem. Theor. Comp.*, 7:3635–3642, 2011.

[7] Kwok-Yan Chan, Leonardo G. Trabuco, Eduard Schreiner, and Klaus Schulten. Cryo-electron microscopy modeling by the molecular dynamics flexible fitting method. *Biopolymers*, 97:678–686, 2012.

[8] Wen Li, Gemma C. Atkinson, Nehal S. Thakor, Ülar Allas, Chuao-chao Lu, Kwok-Yan Chan, Tanel Tenson, Klaus Schulten, Kevin S. Wilson, Vasili Hauryliuk, and Joachim Frank. Mechanism of tetracycline resistance by ribosomal protection protein Tet(O). *Nat. Commun.*, 4:1477, 2013.

[9] Julian Deeng, Kwok-Yan Chan, Otto Berninghausen, Wei Han, James Gumbart, Klaus Schulten, Birgitta Beatrix, and Roland Beckmann. Dynamic behavior of trigger factor on the ribosome. 2014. Submitted.

[10] David B. Wells, Volha Abramkina, and Aleksei Aksimentiev. Exploring transmembrane transport through α-hemolysin with grid-steered molecular dynamics. *J. Chem. Phys.*, 127:125101, 2007.

[11] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. *J. Comp. Chem.*, 26:1781–1802, 2005.

[12] James Gumbart, Leonardo G. Trabuco, Eduard Schreiner, Elizabeth Villa, and Klaus Schulten. Regulation of the protein-conducting channel by a bound ribosome. *Structure*, 17:1453–1464, 2009.

[13] Birgit Seidelt, C. Axel Innis, Daniel N. Wilson, Marco Gartmann, Jean-Paul Armache, Elizabeth Villa, Leonardo G. Trabuco, Thomas Becker, Thorsten Mielke, Klaus Schulten, Thomas A. Steitz, and Roland Beckmann. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science*, 326:1412–1415, 2009.

[14] Thomas Becker, Shashi Bhushan, Alexander Jarasch, Jean-Paul Armache, Soledad Funes, Fabrice Jossinet, James Gumbart, Thorsten Mielke, Otto Berninghausen, Klaus Schulten, Eric Westhof, Reid Gilmore, Elisabet C. Mandon, and Roland Beckmann. Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science*, 326:1369–1373, 2009.

[15] Leonardo G. Trabuco, Christopher B Harrison, Eduard Schreiner, and Klaus Schulten. Recognition of the regulatory nascent chain TnaC by the ribosome. *Structure*, 18:627–637, 2010.

[16] Leonardo G. Trabuco, Eduard Schreiner, John Eargle, Peter Cornish, Taekjip Ha, Zaida Luthey-Schulten, and Klaus Schulten. The role of L1 stalk-tRNA interaction in the ribosome elongation cycle. *J. Mol. Biol.*, 402:741–760, 2010.

[17] Leonardo G. Trabuco, Eduard Schreiner, James Gumbart, Jen Hsin, Elizabeth Villa, and Klaus Schulten. Applications of the molecular dynamics flexible fitting method. *J. Struct. Biol.*, 173:420–427, 2011.

[18] James Gumbart, Eduard Schreiner, Leonardo G. Trabuco, Kwok-Yan Chan, and Klaus Schulten. Viewing the mechanisms of translation through the computational microscope. In Joachim Frank, editor, *Molecular Machines in Biology*, chapter 8, pages 142–157. Cambridge University Press, 2011.

[19] Wen Li, Leonardo G. Trabuco, Klaus Schulten, and Joachim Frank. Molecular dynamics of EF-G during translocation. *Proteins: Struct., Func., Bioinf.*, 79:1478–1486, 2011.

[20] Xabier Agirrezabala, Eduard Schreiner, Leonardo G. Trabuco, Jianlin Lei, Rodrigo F. Ortiz-Meoz, Klaus Schulten, Rachel Green, and Joachim Frank. Structural insights into cognate vs. near-cognate discrimination during decoding. *EMBO J.*, 30:1497–1507, 2011.

[21] Jens Frauenfeld, James Gumbart, Eli O. van der Sluis, Soledad Funes, Marco Gartmann, Birgitta Beatrix, Thorsten Mielke, Otto Berninghausen, Thomas Becker, Klaus Schulten, and Roland Beckmann. Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. *Nat. Struct. Mol. Biol.*, 18:614–621, 2011.

[22] James Gumbart, Christophe Chipot, and Klaus Schulten. Free-energy cost for translocon-assisted insertion of membrane proteins. *Proc. Natl. Acad. Sci. USA*, 108:3596–3601, 2011.

[23] James Gumbart, Christophe Chipot, and Klaus Schulten. Free energy of nascent-chain folding in the translocon. *J. Am. Chem. Soc.*, 133:7602–7607, 2011.

[24] Xabier Agirrezabala, Hstau Liao, Eduard Schreiner, Jie Fu, Rodrigo Ortiz-Meoz, Klaus Schulten, Rachel Green, and Joachim Frank. Structural characterization of mRNA-tRNA translocation intermediates. *Proc. Natl. Acad. Sci. USA*, 109:6094–6099, 2012.

[25] James Gumbart, Eduard Schreiner, Daniel N. Wilson, Roland Beckmann, and Klaus Schulten. Mechanisms of SecM-mediated stalling in the ribosome. *Biophys. J.*, 103:331–341, 2012.

[26] Florence Tama, Osamu Miyashita, and Charles L. Brooks III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J. Struct. Biol.*, 147:315–326, 2004.

[27] Steven J. Ludtke, Phiilip R. Baldwin, and Wah Chiu. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.*, 128:82–97, 1999.

[28] Edward H. Egelman. A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy*, 85:225–234, 2000.

[29] Marin van Heel, George Harauz, Elena V. Oriova, Ralf Schmidt, and Michael Schatz. A new generation of the IMAGIC image processing system. *J. Struct. Biol.*, 116:17–24, 1996.

[30] Xavier Daura, Alan E. Mark, and Wilfred F. van Gunsteren. Peptide folding simulations: no solvent required? *Comput. Phys. Commun.*, 123:97–102, 1999.

[31] Isabella Daidone, Martin B. Ulmschneider, Alfredo Di Nola, Andrea Amadei, and Jeremy C. Smith. Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding. *Proc. Natl. Acad. Sci. USA*, 104:15230–15235, 2007.

[32] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.

[33] David E. Tanner, James C. Phillips, and Klaus Schulten. GPU/CPU algorithm for generalized Born / solvent-accessible surface area implicit solvent calculations. *J. Chem. Theor. Comp.*, 8:2521–2530, 2012.

[34] Steven J. Ludtke, Catherine L. Lawson, Gerard J. Kleywegt, Helen Berman, and Wah Chiu. The 2010 cryo-EM modeling challenge. *Biopolymers*, 97:651–654, 2012.

[35] I. Chopra and M. Roberts. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.*, 65:232–260, 2001.

[36] V. Burdett. Tet(m)-promoted release of tetracycline from ribosomes is GTP dependent. *J. Bacteriol.*, 178:3246–3251, 1996.

[37] S. R. Connell, C. A. Trieber, G. P. Dinos, E. Einfeldt, D. E. Taylor, and K. H. Nierhaus. Mechanism of Tet(O)-mediated tetracycline resistance. *EMBO J.*, 22:945–953, 2003.

[38] Anja Hoffmann, Bernd Bukau, and Günter Kramer. Structure and function of the molecular chaperone trigger factor. *Biochim. Biophys. Acta*, 1803:650–661, 2010.

[39] Anja Hoffmann, Frieder Merz, Anna Rutkowska, Beate Zachmann-Brand, Elke Deuerling, and Bernd Bukau. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J. Biol. Chem.*, 281:6539–6545, 2006.

[40] Vishwas R. Agashe, Suranjana Guha1, Hung-Chun Chang, Pierre Genevaux, Manajit Hayer-Hartl, Markus Stemp, Costa Georgopoulos, F. Ulrich Hartl, and José M.Barral. Function of trigger factor and DnaK in multidomain protein folding. *Cell*, 117:199–209, 2004.

[41] Anja Hoffmann, Annemarie H. Becker, Beate Zachmann-Brand, Elke Deuerling, Bernd Bukau, and Günter Kramer. Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol. Cell*, 2012. http://dx.doi.org/10.1016/j.molcel.2012.07.018.

[42] Elke Deuerling, Agnes Schulze-Specking, Toshifumi Tomoyasu, Axel Mogk, and Bernd Bukau. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature*, 400:693–696, 1999.

[43] Sarah A. Teter, Walid A. Houry, Debbie Ang, Thomas Tradler, David Rockabrand, Gunter Fischer, Paul Blum, Costa Georgopoulos, and F. Ulrich Hartl. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell*, 97:755–765, 1999.

[44] P. Qian, C. N. Hunter, and P. A. Bullough. The 8.5 Å projection structure of the core RC-LH1-PufX dimer of *Rhodobacter sphaeroides*. *J. Mol. Biol.*, 349:948–960, 2005.

[45] X. Agirrezabala, J. Lei, J. L. Brunelle, R. F. Ortiz-Meoz, R. Green, and J. Frank. Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol. Cell*, 32:190–197, 2008.

[46] X. Zhang, L. Jin, Q. Fang, W. H. Hui, and Z. H. Zhou. 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell*, 141:472–482, 2010.

[47] W. Wriggers and P. Chacón. Modeling tricks and fitting techniques for multiresolution structures. *Structure*, 9:779–788, 2001.

[48] N. Volkmann, D. Hanein, G. Ouyang, K. M. Trybus, D. J. DeRosier, and S. Lowey. Evidence for cleft closure in actomyosin upon ADP release. *Nat. Struct. Biol.*, 7:1147–1155, 2000.

[49] M. S. Chapman. Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron-density function. *Acta Cryst. A*, 51:69–80, 1995.

[50] L. F. Chen, E. Blanc, M. S. Chapman, and K. A. Taylor. Real space refinement of acto-myosin structures from sectioned muscle. *J. Struct. Biol.*, 133:221–232, 2001.

[51] W. Wriggers, R. A. Milligan, and J. A. McCammon. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.*, 125:185–195, 1999.

[52] Florence Tama, Osamu Miyashita, and Charles L. Brooks III. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.*, 337:985–999, 2004.

[53] K. Suhre, J. Navaza, and Y. H. Sanejouand. NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Cryst. D*, 62:1098–1100, 2006.

[54] M. Topf, M. L. Baker, M. A. Marc-Renom, W. Chiu, and A. Sali. Refinement of protein structures by iterative comparative modeling and cryoEM density fitting. *J. Mol. Biol.*, 357:1655–1668, 2006.

[55] G. F. Schröder, A. T. Brunger, and M. Levitt. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure*, 15:1630–1641, 2007.

[56] J. Zhu, L. Cheng, Q. Fang, Z. H. Zhou, and B. Honig. Building and refining protein models within cryo-electron microscopy density maps based on homology modeling and multiscale structure refinement. *J. Mol. Biol.*, 397:835–851, 2010.

[57] J. M. Hogle, M. Chow, and D. J. Filman. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science*, 229:1358, 1985.

[58] D. A. Doyle, J. M. Cabral, R. A. Pfuetzer, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon. The structure of the potassium channel: Molecular basis of K^+ conduction and selectivity. *Science*, 280:69–77, 1998.

[59] Peter G. Wolynes. Symmetry and the energy landscapes of biomolecules. *Proc. Natl. Acad. Sci. USA*, 93:14249–14255, 1996.

[60] Trinh Xuan Hoang, Antonio Trovato, Flavio Seno, Jayanth R. Banavar, and Amos Martian. Geometry and symmetry presculpt the free-energy landscape of proteins. *Proc. Natl. Acad. Sci. USA*, 101:7960–7964, 2004.

[61] Ingemar André, Charlie E. M. Strauss, David B. Kaplan, Phillp Bradley, and David Baker. Emergence of symmetry in homooligomeric biological assemblies. *Proc. Natl. Acad. Sci. USA*, 105:16148–16152, 2008.

[62] David S. Goodsell and Arthur J. Olson. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.*, 29:105–153, 2000.

[63] Tom L. Blundell and N. Srinivasan. Symmetry. stability and dynamics of multidomain and multicomponent protein systems. *Proc. Natl. Acad. Sci. USA*, 93:14243–14248, 1996.

[64] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254:1598–1603, 1991.

[65] Hugh Nymeyer, Angel E. García, and José Nelson Onuchic. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA*, 95:5921–5928, 1998.

[66] Ingemar André, Philip Bradley, Chu Wang, and David Baker. Prediction of the structure of symmetrical protein assemblies. *Proc. Natl. Acad. Sci. USA*, 104:17656–17661, 2007.

[67] Andriy Anishkin, Adina L. Milac, and H. Robert Guy. Symmetry-restrained molecular dynamics simulations improve homology models of potassium channels. *Proteins: Struct., Func., Bioinf.*, 78:932–949, 2010.

[68] A. M. Roseman, S. Chen, H. White, K. Braig, and H. R. Saibil. The chaperonin ATPase cycle: Mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell*, 87:241–251, 1996.

[69] Charu Chaudhry, George W. Farr, Matthew J. Todd, Hays S. Rye, Axel T. Brunger, Paul D. Ad ams, Arthur L. Horwich, and Paul B. Sigier. Role of the γ-phosphate of ATP in triggering protein folding by GroEL-GroES: function, structure and energetics. *EMBO J.*, 22:4877–4887, 2003.

[70] Z. Xu, A. L. Horwich, and P. B. Sigler. The crystal structure of the asymmetric groel-groes-$(adp)_7$ chaperonin complex. *Nature*, 388:741–750, 1997.

[71] Neil A. Ranson, Daniel K. Clare, George W. Farr, David Houldershaw, Arthur L. Horwich, and Helen R. Saibil. Allosteric signalling of ATP hydrolysis in GroEL-GroES complexes. *Nat. Struct. Mol. Biol.*, 13:147–152, 2006.

[72] Dakshina Jandhyala, Mark Berman, Paul R. Meyers, B. Trevor Sewell, Richard C. Wilson, and Mic hael J. Benedik. CynD, the Cyanide Dihydratase from *Bacillus pumilus*: Gene cloning and structural studies. *Appl. Environ. Microbiol.*, 69:4794–4805, 2003.

[73] R. Ndoria Thuku, D Brady, Michael J. Benedik, and B. Trevor Sewell. Microbial nitrilases: versatile, spiral forming, industrial enzymes. *J. Appl. Microbiol.*, 106:703–727, 2009.

[74] R. Ndoria Thuku, Brandon W. Weber, Arvind Varsani, and B. Trevor Sewell. Post-translational cleavage of recombinantly expressed nitrilase from *Rhodococcus rhodochrous* J1 yields a stable, active helical form. *FEBS J.*, 274:2099–2108, 2007.

[75] J.D. Woodwarda, B.W. Webera, M.P. Scheffera, M.J. Benedikc, A. Hoengerd, and B.T. Sewell. Helical structure of unidirectionally shadowed metal replicas of cyanide hydratase from *Gloeocercospora sorghi*. *J. Struct. Biol.*, 161:111–119, 2008.

[76] Junjie Zhang, Matthew L. Baker, Gunnar F. Schröder, Nicholai R. Douglas, Stefanie Reissmann, Joanita Jakana, Matthew Dougherty, Caroline J. Fu, Michael Levitt, Steven J. Ludtke, Judith Frydman, and Wah Chiu. Mechanism of folding chamber closure in a group ii chaperonin. *Nature*, 463:379–383, 2010.

[77] Jose H. Pereira, Corie Y. Ralston, Nicholai R. Douglas, Daniel Meyer, Kelly M. Knee, Daniel R. Goulet, Jonathan A. King, Judith Frydman, and Paul D. Adams. Crystal structures of a group II chaperonin reveal the open and closed states associated with the protein folding cycle. *J. Biol. Chem.*, 285:27958–27966, 2010.

[78] James M. Hogle. Poliovirus cell entry: Common structural themes in viral cell entry pathways. *Annu. Rev. Microbiol.*, 56:677–702, 2002.

[79] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[80] Alexander D. MacKerell, Jr., Michael Feig, and Charles L. Brooks, III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.*, 25(11):1400–1415, 2004.

[81] Helmut Grubmüller, Helmut Heller, Andreas Windemuth, and Klaus Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Sim.*, 6:121–142, 1991.

[82] M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97:1990–2001, 1992.

[83] Eric H. Lee, Jen Hsin, Marcos Sotomayor, Gemma Comellas, and Klaus Schulten. Discovery through the computational microscope. *Structure*, 17:1295–1306, 2009.

[84] Peter L. Freddolino and Klaus Schulten. Common structural transitions in explicit-solvent simulations of villin headpiece folding. *Biophys. J.*, 97:2338–2347, 2009.

[85] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. Interaction models for water in relation to protein hydration. In B. Pullman, editor, *Intermolecular Forces*, pages 331–342. D. Reidel Publishing Company, 1981.

[86] W. Clark Still, Anna Tempczyk, Ronald C. Hawley, and Thomas Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[87] S. A. Hassan and E. L. Mehler. A critical analysis of continuum electrostatics: the screened Coulomb potential-implicit solvent model and the study of the alanine dipeptide and discrimination of misfolded structures of proteins. *Proteins: Struct., Func., Gen.*, 47:45–61, 2002.

[88] Michael Holst, Nathan Baker, and F. Wang. Adaptive multilevel finite element solution of the Poisson–Boltzmann equation II. refinement at solvent-accessible surfaces in biomolecular systems. *J. Comp. Chem.*, 21:1343–1352, 2000.

[89] Nilashis Nandi and Biman Bagchi. Dielectric relaxation of biological water. *J. Phys. Chem. B*, 101(50):10954–10961, 1997.

[90] Young Min Rhee and Vijay S. Pande. Solvent viscosity dependence of the protein folding dynamics. *J. Phys. Chem. B*, 112:6221–6227, 2008.

[91] Benzhuo Lu, Xiaolin Cheng, Jingfang Huang, and J. Andrew McCammon. AFMPB: An adaptive fast multipole Poisson-Boltzmann solver for calculating electrostatics in biomolecular systems. *Comput. Phys. Commun.*, 181:1150–1160, 2010.

[92] N. A. Baker. Improving implicit solvent simulations: a Poisson-centric view. *Curr. Opin. Struct. Biol.*, 15:137–143, 2005.

[93] S. A. Hassan, E. L. Mehler, D. Zhang, and H. Weinstein. Molecular dynamics simulations of peptides and proteins with a continuum electrostatic model based on screened Coulomb potentials. *Proteins: Struct., Func., Gen.*, 51:109–125, 2003.

[94] Michael Schaefer and Martin Karplus. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.*, 100:1578–1599, 1996.

[95] Di Qiu, Peter S. Shenkin, Frank P. Hollinger, and W. Clark Still. The gb/sa continuum model for solvation. a fast analytical method for the calculation of approximate born radii. *J. Phys. Chem.*, 101:3005–3014, 1997.

[96] Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4:187–217, 1983.

[97] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. Charmm: The biomolecular simulation program. *J. Comp. Chem.*, 30:1545–1614, 2009.

[98] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theor. Comp.*, 4:435–447, 2008.

[99] Per Larsson and Erik Lindahl. A high-performance parallel-generalized born implementation enabled by tabulated interaction rescaling. *J. Comp. Chem.*, 31:2593–2600, 2010.

[100] David A. Pearlman, David A. Case, James W. Caldwell, Wilson S. Ross, Thomas E. Cheatham, Steve DeBolt, David Ferguson, George Seibel, and Peter Kollman. Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, 91:1–41, 1995.

[101] David E. Tanner, Wen Ma, Zhongzhou Chen, and Klaus Schulten. Theoretical and computational investigation of flagellin translocation and bacterial flagellum growth. *Biophys. J.*, 100:2548–2556, 2011.

[102] Michael Feig, Alexey Onufreiv, Michael S. Lee, Wonpil Im, David A. Case, and Charles L. Brooks. Performance comparison of generalized born and poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comp. Chem.*, 25:265–284, 2004.

[103] Michael S. Lee, Freddie R. Salsbury, and Charles L. Brooks. A high-performance parallel-generalized born implementation enabled by tabulated interaction rescaling. *J. Chem. Phys.*, 116:10606–10614, 2002.

[104] Wonpil Im, Michael S. Lee, and Charles L. Brooks. Generalized born model with a simple smoothing function. *J. Comp. Chem.*, 24:1691–1702, 2003.

[105] Devleena Shivakumar, Yuqing Deng, and Benoit Roux. Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model. *J. Chem. Theor. Comp.*, 5:919–930, 2009.

[106] Barry J. Grant, Alemayehu A. Gorfe, and J. Andrew McCammon. Large conformational changes in proteins: signaling and other functions. *Theoret. Chim. Acta*, 20:142–147, 2010.

[107] Roland Schulz, Benjamin Lindner, Loukas Petridis, and Jeremy C. Smith. Scaling of multimillion-atom biological molecular dynamics simulation on a petascale supercomputer. *J. Chem. Theor. Comp.*, 5:2798–2808, 2009.

[108] P. V. Cornish, D. N. Ermolenko, H. F. Noller, and T. Ha. Spontaneous intersubunit rotation in single ribosomes. *Mol. Cell*, 30:578–588, 2008.

[109] P. Cornish, D. N. Ermolenko, D. W. Staple, L. Hoang, R. P. Hickerson, H. F. Noller, and T. Ha. Following movement of the L1 stalk between three functional states in single ribosomes. *Proc. Natl. Acad. Sci. USA*, 106:2571–2576, 2009.

[110] L. H. Horan and H. F. Noller. Intersubunit movement is required for ribosomal translocation. *Proc. Natl. Acad. Sci. USA*, 104:4881–4885, 2007.

[111] L. Stella and S. Melchionna. Equilibration and sampling in molecular dynamics simulations of biomolecules. *J. Chem. Phys.*, 109:10115–10117, 1998.

[112] Iwona K. Wower, Jacek Wower, and Robert A. Zimmermann. Ribosomal protein L27 participates in both 50S subunit assembly and the peptidyl transferase reaction. *J. Biol. Chem.*, 273:19847–19852, 1998.

[113] A. R. Cukras, D. R. Southworth, J. L. Brunelle, G. M. Culver, and R. Green. Ribosomal proteins S12 and S13 function as control elements for translocation of the mRNA:tRNA complex. *Mol. Cell*, 12:321–328, 2003.

[114] M. Diaconu, U. Kothe, F. Schlünzen, N. Fischer, J. M. Harms, A. G. Tonevitsky, H. Stark, M. V. Rodnina, and M. C. Wahl. Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell*, 121:991–1004, 2005.

[115] V. Berk, W. Zhang, R. D. Pai, and J. H. D. Cate. Structural basis for mRNA and tRNA positioning on the ribosome. *Proc. Natl. Acad. Sci. USA*, 103:15830–15834, 2006.

[116] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[117] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[118] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65:712–725, 2006.

[119] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.*, 92:3817–3829, 2007.

[120] R. Aduri, B. T. Psciuk, P. Saro, H. Taniga, H. B. Schlegel, and J. SantaLucia. AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J. Chem. Theor. Comp.*, 3:1464–1475, 2007.

[121] Joachim Frank. Single-particle reconstruction of biological macromolecules in electron microscopy - 30 years. *Quart. Rev. Biophys.*, 42:139–158, 2009.

[122] Steven J. Ludtke, Matthew L. Baker, Dong-Hua Chen, Jiu-Li Song, David T. Chuang, and Wah Chiu. De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure*, 16:441–448, 2008.

[123] Xing Zhang, Ethan Settembre, Chen Xu, Philip R. Dormitzer, Richard Bellamy, Stephen C. Harrison, and Nikolaus Grigorieff. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. USA*, 105:1867–1872, 2008.

[124] Corey F. Hryc, Dong-Hua Chen, and Wah Chiu. Near-atomic-resolution cryo-EM for molecular virology. *Curr. Opin. Virol.*, 1:110–117, 2011.

[125] Matthew L. Baker, Junjie Zhang, Steven J. Ludtke, and Wah Chiu. Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nat. Protoc.*, 5:1697–1708, 2010.

[126] Rui Zhang, Corey F. Hryc, Yao Cong, Xiangan Liu, Joanita Jakana, Rodion Gorchakov, Matthew L. Baker, Scott C. Weaver, and Wah Chiu. 4.4 å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. *EMBO J.*, 30:3854–3863, 2011.

[127] Saori Maki-Yonekura, Koji Yonekura, and Keiichi Namba. Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. *Nat. Struct. Mol. Biol.*, 17:417–422, 2010.

[128] Lingpeng Cheng, Jiang Zhu, Wong Hoi Hui, Xiaokang Zhang, Barry Honig, Qin Fang, and Z. Hong Zhou. Backbone model of an aquareovirus virion by cryo-electron microscopy and bioinformatics. *J. Mol. Biol.*, 397:852–863, 2010.

[129] Eduard Schreiner, Leonardo G. Trabuco, Peter L. Freddolino, and Klaus Schulten. Stereochemical errors and their implications for molecular dynamics simulations. *BMC Bioinform*, 12:190, 2011.

[130] Rhiju Das and David Baker. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.*, 77:363–382, 2008.

[131] Matthew L. Baker, Mariah R. Baker, Corey F. Hryc, and Frank DiMaio. Analyses of sub-nanometer resolution cryo-EM density maps. *Meth. Enzym.*, 483:1–29, 2010.

[132] C. C. Jolley, S. A. Wells, P. Fromme, and M. F. Thorpe. Fitting low-resolution cryo-EM maps of proteins using constrained geometric simulations. *Biophys. J.*, 94:1613–1621, 2008.

[133] Daniel W. Farrell, Kirill Speranskiy, and M. F. Thorpe. Generating stereochemically acceptable protein pathways. *Proteins: Struct., Func., Bioinf.*, 78:2908–2921, 2010.

[134] Vincent B. Chen, W. Bryan Arendall III, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst. D*, 66:12–21, 2010.

[135] Richard K. Hite, Zongli Li, and Thomas Walz. Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. *EMBO J.*, 29:1652–1658, 2010.

[136] Ryan McGreevy, Abhishek Singharoy, Qufei Li, Jingfen Zhang, Eduardo Perozo, and Klaus Schulten. xMDFF: Molecular dynamics flexible fitting of low resolution X-Ray structures. 2014. Submitted.

[137] Melih K. Sener, Jen Hsin, Leonardo G. Trabuco, Elizabeth Villa, Pu Qian, C. Neil Hunter, and Klaus Schulten. Structural model and excitonic properties of the dimeric RC-LH1-PufX complex from Rhodobacter sphaeroides. *Chem. Phys.*, 357:188–197, 2009.

[138] Jen Hsin, James Gumbart, Leonardo G. Trabuco, Elizabeth Villa, Pu Qian, C. Neil Hunter, and Klaus Schulten. Protein-induced membrane curvature investigated through molecular dynamics flexible fitting. *Biophys. J.*, 97:321–329, 2009.

[139] Philip D. Kiser, George H. Lorimer, and Krzysztof Palczewski. Use of thallium to identify monovalent cation binding sites in GroEL. *Acta. Cryst. F*, 65:967–971, 2009.

[140] Andrej Sali and Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, 234:779, 1993.

[141] Magali Mathieu, Isabelle Petitpas, Jorge Navaza, Jean Lepault, Evelyne Kohli, Pierre Pothier, B.V.Venkataram Prasad, Jean Cohen, and Félix A. Rey. Atomic structure of the major capsid protein of rotavirus: implications for the architecture of the virion. *EMBO J.*, 20:1485–1497, 2001.

[142] R. M. Voorhees, A. Weixlbaumer, D. Loakes, A. C. Kelley, and V. Ramakrishnan. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat. Struct. Mol. Biol.*, 16:528–533, 2009.

[143] Jan-Christian Schuette, Frank V Murphy, Ann C Kelley, John R Weir, Jan Giesebrecht, Sean R Connell, Justus Loerke, Thorsten Mielke, Wei Zhang, Pawel A Penczek, V Ramakrishnan, and Christian M T Spahn. Gtpase activation of elongation factor ef-tu by the ribosome during decoding. *EMBO J.*, 28(6):755–65, 2009.

[144] D. E. Brodersen, W. M. Clemons, A. P. Carter, R. J. Morgan-Warren, B. T. Wimberly, and V. Ramakrishnan. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell*, 103:1143–1154, 2000.

[145] B. Epe, P. Woolley, and H. Hornig. Competition between tetracycline and trna at both p and a sites of the ribosome of *Escherichia coli*. *FEBS Lett.*, 213:443–447, 1987.

[146] S. R. Connell, D. M. Tracz, K. H. Nierhaus, and D. E. Taylor. Ribosomal protection proteins and their mechanism of tetracycline resistance. *Antimicrob. Agents Chemother.*, 47:3675–3681, 2003.

[147] M. C. Roberts. Update on acquired tetracycline resistance genes. *FEMS Microbiol. Lett.*, 245:195–203, 2005.

[148] T. Margus, M. Remm, and T. Tenson. Phylogenetic distribution of translational GTPases in bacteria. *BMC Genomics*, 8:15, 2007.

[149] C. M. Spahn, G. Blaha, R. K. Agrawal, P. Penczek, R. A. Grassucci, C. A. Trieber, S. R. Connell, D. E. Taylor, K. H. Nierhaus, and J. Frank. Localization of the ribosomal protection protein Tet(O) on the ribosome and the mechanism of tetracycline resistance. *Mol. Cell*, 7:1037–1045, 2001.

[150] N. S. Thakor, R. Nechifor, P. G. Scott, M. Keelan, D. E. Taylor, and K. S. Wilson. Chimeras of bacterial translation factors Tet(O) and EF-G. *FEBS Lett.*, 582:1386–1390, 2008.

[151] R. Sanchez-Pescador, J. T. Brown, M. Roberts, and M. S. Urdea. Homology of the TetM with translational elongation factors: implications for potential modes of tetm-conferred tetracycline resistance. *Nucleic Acids Res.*, 16:1218, 1988.

[152] Y.-G. Gao, M. Selmer, C. M. Dunham, A. Weixlbaumer, A. C. Kelley, and V. Ramakrishnan. The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science*, 326:694–699, 2009.

[153] O. O'Sullivan, K. Suhre, C. Abergel, D. G. Higgins, and C. Notredame. 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *J. Mol. Biol.*, 340:385–395, 2004.

[154] G. C. Atkinson and S. L. Baldauf. Evolution of elongation factor G and the origins of mitochondrial and chloroplast forms. *Mol. Biol. Evol.*, 28:1281–1292, 2011.

[155] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, and A Sali. *Comparative protein structure modeling using Modeller*, chapter 5, Unit 5.6. John Wiley & Sons, Inc., 2006.

[156] M. Selmer, C. M. Dunham, F. V. Murphy IV, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir, and V. Ramakrishnan. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, 313:1935–1942, 2006.

[157] J. M. Ogle, D. E. Brodersen amd W. M. Clemons, M. J. Tarry, A. P. Carter, and V. Ramakrishnan. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science*, 292:897–902, 2001.

[158] Rebecca M. Voorhees, T. Martin Schmeing, Ann C. Kelley, and V. Ramakrishnan. The mechanism for activation of GTP hydrolysis on the ribosome. *Science*, 330:835–838, 2010.

[159] S. R. Connell, C. A. Trieber, U. Stelzl, E. Einfeldt, D. E. Taylor, and K. H. Nierhaus. The tetracycline resistance protein Tet(o) perturbs the conformation of the ribosomal decoding centre. *Mol. Microbiol.*, 45:1463–1472, 2002.

[160] C. A. Trieber, N. Burkhardt, K. H. Nierhaus, and D. E. Taylor. Ribosomal protection from tetracycline mediated by Tet(O): Tet(O) interaction with ribosomes is GTP-dependent. *Biol. Chem.*, 379:847–855, 1998.

[161] J. Fei, P. Kosuri, D. D. MacDougall, and R. L. Gonzalez, Jr. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. *Mol. Cell*, 30:348–359, 2008.

[162] D. J. Taylor, J. Nilsson, A. R. Merrill, G. R. Andersen, P. Nissen, and J. Frank. Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J.*, 26:2421–2431, 2007.

[163] A. Dönhöfer, S. Franckenberg, S. Wickles, O. Berninghausen, R. Beckmann, and D. N. Wilson. Structural basis for TetM-mediated tetracycline resistance. *Proc. Natl. Acad. Sci. USA*, 109:16900–16905, 2012.

[164] K. Katoh, K. Kuma, H. Toh, and T. Miyata. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.*, 33:511–518, 2005.

[165] M. N. Price, P. S. Dehal, and A. P. Arkin. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One*, 5:e9490, 2010.

[166] M. Laurberg, O. Kristensen, K. Martemyanov, A. T. Gudkov, I. Nagaev, D. Hughes, and A. Liljas. Structure of a mutant EF-G reveals domain III and possibly the fusidic acid binding site. *J. Mol. Biol.*, 303:593–603, 2000.

[167] Lars Ferbitz, Timm Maier, Holger Patzelt, Bernd Bukau, Elke Deuerling, and Nenad Ban. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*, 431:590–596, 2004.

[168] Pierre Genevaux, France Keppel, Franoise Schwager, Petra S. LangendijkGenevaux, F. Ulrich Hartl, and Costa Georgopoulos. *In vivo* analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep.*, 5:195–200, 2004.

[169] Frieder Merz, Anja Hoffmann, Anna Rutkowska, Beate Zachmann-Brand, Bernd Bukau, and Elke Deuerling. The c-terminal domain of *Escherichia coli* trigger factor represents the central module of its chaperone activity. *J. Biol. Chem.*, 281:31963–31971, 2006.

[170] Rashmi Gupta, Sathish Kumar Lakshmipathy, Hung-Chun Chang, Stephanie A. Etchells, and F. Ulrich Hartl. Trigger factor lacking the PPIase domain can enhance the folding of eukaryotic multi-domain proteins in *Escherichia coli*. *FEBS Lett.*, 584:3620–3624, 2010.

[171] Sathish Kumar Lakshmipathy, Rashmi Gupta, Stefan Pinkert, Stephanie Anne Etchells, and F. Ulrich Hartl. Versatility of trigger factor interactions with ribosome-nascent chain complexes. *J. Biol. Chem.*, 285:27911–27923, 2010.

[172] S. Bhushan, M. Gartmann, M. Halic, J. P. Armache, A. Jarasch, T. Mielke, O. Berninghausen, D. N. Wilson, and R. Beckmann. α-helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.*, 17:313–317, 2010.

[173] J. Lu and C. Deutsch. Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.*, 12:1123–1129, 2005.

[174] C. A. Woolhead, P. J. McCormick, and A. E. Johnson. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell*, 116:725–736, 2004.

[175] Daniel N. Wilson and Roland Beckmann. The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Curr. Opin. Struct. Biol.*, 21:274–282, 2011.

[176] David Baram, Erez Pyetan, Assa Sittner, Tamar Auerbach-Nevo, Anat Bashan, and Ada Yonath. Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proc. Natl. Acad. Sci. USA*, 102:1201712022, 2005.

[177] Frank Schlünzen, Daniel N. Wilson, Pingsheng Tian, Jörg M. Harms, Stuart J. McInnes, Harly A. S. Hansen, Renate Albrecht, Jörg Buerger, Sigurd M. Wilbanks, and Paola Fucini. The binding mode of the trigger factor on the ribosome: Implications for protein folding and SRP interaction. *Structure*, 13:16851694, 2005.

[178] Frieder Merz, Daniel Boehringer, Christiane Schaffitzel, Steffen Preissler, Anja Hoffmann, Timm Maier, Anna Rutkowska, Jasmin Lozza, Nenad Ban, Bernd Bukau, and Elke Deuerling. Molecular mechanism and structure of trigger factor bound to the translating ribosome. *EMBO J.*, 27:1622–1632, 2008.

[179] Eugene Oh, Annemarie H. Becker, Arzu Sandikci, Damon Huber, Rachna Chaba, Felix Gloge, Robert J. Nichols, Athanasios Typas, Carol A. Gross, Günter Kramer, Jonathan S. Weissman, and Bernd Bukau. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor *In Vivo*. *Cell*, 147:12951308, 2011.

[180] Ole Kristensen and Michael Gajhede. Chaperone binding at the ribosomal exit tunnel. *Structure*, 11:15471556, 2003.

[181] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of Coot. *Acta Cryst. D*, 66:486–501, 2010.

[182] Wei Han, Cheuk-Kin Wan, and Yun-Dong Wu. Toward a coarse-grained model coupled with a coarse-grained solvent model: solvation free energies of amino acid side chain. *J. Chem. Theor. Comp.*, 4:1891–1901, 2008.

[183] Wei Han, Cheuk-Kin Wan, Fan Jiang, and Yun-Dong Wu. PACE force field for protein simulations. 1. full parameterization of version 1 and verification. *J. Chem. Theor. Comp.*, 6:3373–3389, 2010.

[184] Wei Han, Cheuk-Kin Wan, and Yun-Dong Wu. PACE force field for protein simulations. 2. folding simulations of peptides. *J. Chem. Theor. Comp.*, 6:3390–3402, 2010.

[185] Wei Han and Klaus Schulten. Further optimization of a hybrid united-atom and coarse-grained force field for folding simulations: Improved backbone hydration and interactions between charged side chains. *J. Chem. Theor. Comp.*, 8:4413–4424, 2012.

[186] Wei Han and Klaus Schulten. Characterization of folding mechanisms of Trp-cage and WW-domain by network analysis of simulations with a hybrid-resolution model. *J. Phys. Chem. B*, 117:13367–13377, 2013. doi: 10.1021/jp404331d.

[187] A. Bakan, L. M. Meireles, and I. Bahar. ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics*, 27(11):1575–1577, 2011.

[188] Andrew S. Thomas, Suifang Mao, and Adrian H. Elcock. Flexibility of the bacterial chaperone trigger factor in microsecond-timescale molecular dynamics simulations. *Biophys. J.*, 105:732–744, 2013.

[189] Kushagra Singhal, Jocelyne Vreede, Alireza Mashaghi, Sander J. Tans, and Peter G. Bolhuis. Hydrophobic collapse of trigger factor monomer in solution. *PLoS One*, 8:e59683, 2013.

[190] Holger Patzelt, Stefan Rüdiger, Dirk Brehmer, Günter Kramer, Sonja Vorderwülbecke, Elke Schaffitzel, Andreas Waitz, Thomas Hesterkamp, Liying Dong, Jens Schneider-Mergener, Bernd Bukau, and Elke Deuerling. Binding specificity of *Escherichia coli* trigger factor. *Proc. Natl. Acad. Sci. USA*, 98:14244–14249, 2001.

[191] Christian M. Kaiser, Hung-Chun Chang, Vishwas R. Agashe, Sathish K. Lakshmipathy, Stephanie A. Etchells, Manajit Hayer-Hartl, F. Ulrich Hartl, and José M. Barral. Real-time observation of trigger factor function on translating ribosomes. *Nature*, 444:455–460, 2006.

[192] Anna Rutkowska, Matthias P Mayer, Anja Hoffmann, Frieder Merz, Beate Zachmann-Brand, Christiane Schaffitzel, Nenad Ban, Elke Deuerling, and Bernd Bukau. Dynamics of trigger factor interaction with translating ribosomes. *J. Biol. Chem.*, 283:4124–4132, 2008.

[193] Johannes Söding, Andreas Biegert, and Andrei N. Lupas. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, 33:W244–W248, 2005.

[194] John E. Stone, James C. Phillips, Peter L. Freddolino, David J. Hardy, Leonardo G. Trabuco, and Klaus Schulten. Accelerating molecular modeling applications with graphics processors. *J. Comp. Chem.*, 28:2618–2640, 2007.

[195] Cheuk-Kin Wan, Wei Han, and Yun-Dong Wu. Parameterization of PACE force field for membrane environment and simulation of helical peptides and helix-helix association. *J. Chem. Theor. Comp.*, 8(1):300–313, 2012.

[196] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111:7812–7824, 2007.

[197] Syma Khalid, Peter J. Bond, John Holyoake, Robert W. Hawtin, and Mark S. P. Sansom. DNA and lipid bilayers: self-assembly and insertion. *J. R. Soc. Interface*, 124:S241–S250, 2008.

[198] E. P. O'Brien, S. D. Hsu, J. Christodoulou, M. Vendruscolo, and C. M. Dobson. Transient tertiary structure formation within the ribosome exit port. *J. Am. Chem. Soc.*, 132:16928–16937, 2010.

[199] Edward P. OBrien, John Christodoulou, Michele Vendruscolo, and Christopher M. Dobson. New scenarios of protein folding can occur on the ribosome. *J. Am. Chem. Soc.*, 133:513–526, 2011.

[200] Edward P. O'Brien, John Christodoulou, Michele Vendruscolo, and Christopher M. Dobson. Trigger factor slows co-translational folding through kinetic trapping while sterically protecting the nascent chain from aberrant cytosolic interactions. *J. Am. Chem. Soc.*, 134:10920–10932, 2012.

[201] Gongpu Zhao, Juan R. Perilla, Ernest L. Yufenyuy, Xin Meng, Bo Chen, Jiying Ning, Jinwoo Ahn, Angela M. Gronenborn, Klaus Schulten, Christopher Aiken, and Peijun Zhang. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497:643–646, 2013.

[202] Andreas M. Anger, Jean-Paul Armache, Otto Berninghausen, Michael Habeck, Marion Subklewe, Daniel N. Wilson, and Roland Beckmann. Structures of the human and *Drosophila* 80S ribosomes. *Nature*, 497:80–85, 2013.

[203] A. Bondi. van der waals volumes and radii. *J. Phys. Chem.*, 68(3):441–451, 1964.

[204] Jayashree Srinivasan, Megan W. Trevathan, Paul Beroza, and David A. Case. Application of a pairwise generalized born model to proteins and nucleic acids: inclusion of salt effects. *Theoret. Chim. Acta*, 101:426–434, 1999.

[205] Lexey Onufriev, David A. Case, and Donald Bashford. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *J. Comp. Chem.*, 23:1297–1304, 2002.

[206] M. Feig and C. L. Brooks III. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.*, 14:217–224, 2004.

[207] Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, 100:19824–19839, 1996.

[208] Michael Schaefer and Cornelius Froemmel. A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *J. Mol. Biol.*, 216:1045 – 1066, 1990.

[209] Alexey Onufriev, Donald Bashford, and David A. Case. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem.*, 104:3712–3720, 2000.

[210] Alexey Onufriev, Donald Bashford, and David A. Case. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins: Struct., Func., Bioinf.*, 55:383–394, 2004.